

UNIVERSIDADE DE LISBOA
Faculdade de Ciências
Departamento de Informática



**MOVE AND SPEAK: INTERAÇÃO COM MÚLTIPLAS
SUPERFÍCIES ATRAVÉS DE GESTOS E VOZ**

Andreia Sofia Gomes Ribeiro

DISSERTAÇÃO

MESTRADO EM ENGENHARIA INFORMÁTICA
Especialização em Sistemas de Informação

2012

UNIVERSIDADE DE LISBOA
Faculdade de Ciências
Departamento de Informática



**MOVE AND SPEAK: INTERAÇÃO COM MÚLTIPLAS
SUPERFÍCIES ATRAVÉS DE GESTOS E VOZ**

Andreia Sofia Gomes Ribeiro

DISSERTAÇÃO

Dissertação orientada pelo Prof. Doutor Carlos Alberto Pacheco dos Anjos Duarte

MESTRADO EM ENGENHARIA INFORMÁTICA
Especialização em Sistemas de Informação

2012

Agradecimentos

Agradeço ao Professor Carlos Duarte por me ter orientado durante o decorrer do trabalho, pela transmissão de ideias, apoio e motivação. Agradeço também a compreensão, disponibilidade e paciência que teve comigo.

À Faculdade de Ciências, ao LASIGE, ao HCIM e a todos os seus membros, por me terem dado todas as condições necessárias à produção do trabalho.

Aos meus Pais por tudo o que me ensinaram, pelos valores transmitidos e por me terem dado as condições e a oportunidade de estudar até onde desejei.

Ao meu Irmão pelo apoio e carinho que sempre me ofereceu.

Ao meu Namorado por ter sempre uma palavra de apoio na altura certa, pela compreensão, companheirismo, amizade e o amor que demonstrou ao longo destes anos. Amo-te!

A todos os meus amigos e colegas da Faculdade, que me proporcionaram momentos que nunca mais irei esquecer.

Por último gostaria de agradecer ao Rafael Nunes na ajuda do desenvolvimento dos protótipos desenvolvidos.

Aos meus Pais, Helena e Jorge

Resumo

Os gestos e a voz são formas de fornecer interação natural e intuitiva com dispositivos em muitos domínios dos sistemas computacionais. A interação gestual, principalmente, está muito voga na área dos jogos onde o movimento do corpo atua como comando para controlar o jogo. Esta interação foi alargada aos computadores, alterando assim a interação pessoa-máquina.

O objetivo deste trabalho é proceder ao estudo da interação gestual, com e sem ajuda de toque, identificando quais os gestos e comandos de voz que são naturalmente empregues para cada ação. Esta interação será estudada em aplicações onde são permitidos multimodalidades, multiutilizadores e multisuperfícies. A interação é feita em superfícies de tamanho variado, utilizadas em simultâneo. O utilizador interage com uma superfície que se encontre perto dele, onde a interação é feita através de toque e voz, e passa informação desejada para uma superfície de grande dimensão, uma projeção, onde a interação será feita à distância através de gestos e voz.

Neste trabalho explora-se o uso da interação gestual com ou sem utilização de voz, e em particular, o impacto que o número de utilizadores e o funcionamento da aplicação têm na interação. Foram feitos dois estudos: o primeiro identifica quais os gestos e comandos de voz que são adequados a determinada ação; o segundo analisa o impacto que o número de utilizadores e o funcionamento da aplicação têm na utilização de aplicações que suportam este tipo de interação. Os resultados demonstraram que para a construção de aplicações multimodais, multiutilizador e multisuperfície os comandos de voz em adição aos gestos são um requisito essencial a ter em conta, pois desambigam os gestos e tornam mais fácil a manipulação de objetos virtuais para ações abstratas. Verificou-se, também, que o número de pessoas tem influência no número de comandos executados e que o funcionamento da aplicação também é um fator a ter em conta na construção destas aplicações.

Palavras-chave: Gestos, Comandos de Voz, Interação, Mesa Interativa Multi-Toque, Superfície de Projeção, Superfície Multi-Toque, Multimodal, Multiutilizador, Multisuperfície, Aplicações para manipulação de objectos virtuais, Estudos com Utilizadores

Abstract

Gestures and speech are capable of providing natural, intuitive interaction mechanisms for several computer systems domains. Gesture interaction, in particular, is currently trending in computer games where the whole body can act as a game controller. This type of interaction is being extended to traditional computer systems, thus impacting human-machine interaction.

The goal of this work is to study gestural interaction, in scenarios with or without support for touch, identifying which gestures and speech commands are naturally employed by users. The interaction will be studied with an application supporting multi modality, multi users and multi surfaces (including different surfaces which can be used simultaneously). In this application, the user can interact through touch and speech with a close-by surface, and through gestures and speech with a projected surface.

This work explores, in particular, the effect that the number of users and the placement of new and copied objects has on the interaction. Two studies have been conducted. The first one tries to discover which gestures and speech commands are appropriate for different actions. The second one studies the effect of the aforementioned factors on the interaction. Results show that, when building multimodal, multiuser, multi surface applications, speech commands should complement gestures, since they play a vital role in disambiguating gestures, making abstract commands easier to specify. Other factors that must be taken into account are the number of users and how the application handles object copies and creation.

Keywords: Gestures, Speech commands, Interaction, Multitouch interactive table, Projection surface, Multitouch surface, Multimodal, Multiuser, Multi surface, Virtual objects manipulation applications, User studies gestures, voice, interaction, large surfaces

Conteúdo

Lista de Figuras	xvi
Lista de Tabelas	xix
1 Introdução	1
1.1 Projecto VISTA	1
1.2 Motivação	2
1.3 Objetivos	3
1.4 Contribuições	4
1.5 Estrutura do documento	4
2 Trabalho relacionado	7
2.1 Interação Gestual	7
2.1.1 Interação Gestual sem Voz	7
2.1.2 Interação Gestual com Voz	11
2.2 Tipos de Interação	13
2.2.1 Interação com uma superfície de apoio	13
2.2.2 Interação sem uma superfície de apoio	15
2.3 Colaboração/Comunicação	17
2.4 Discussão	18
3 Desenho de Aplicações Multimodais, Multiutilizador e Multisuperfície	21
3.1 Processo de Desenho	21
3.2 Aplicação do Processo de Desenho em Aplicações Multimodais e Multi-superfície	22
3.2.1 Primeira Iteração do Processo de Desenho	23
3.2.2 Segunda Iteração do Processo de Desenho	24
3.3 Componentes (Hardware)	25
3.4 Tecnologias	26
3.5 Arquitetura	27
3.6 Funcionalidades	29

4	Estudos de Interação	33
4.1	Primeiro Estudo	33
4.1.1	Enquadramento e preparação do estudo	33
4.1.2	Tarefas	34
4.1.3	Participantes	34
4.1.4	Procedimento	35
4.1.5	Resultados	35
4.1.6	Análise e Discussão dos resultados	36
4.2	Segundo Estudo	42
4.2.1	Enquadramento e preparação do estudo	42
4.2.2	Tarefas	42
4.2.3	Participantes	44
4.2.4	Procedimento	44
4.2.5	Análise dos Resultados	47
4.2.6	Discussão dos resultados	54
5	Conclusão	57
5.1	Conclusão	57
5.2	Trabalho Futuro	59
A	Questionário	61
	Bibliografia	69
	Índice	70

Lista de Figuras

2.1	Diferentes formas de ampliar [26].	9
2.2	Gestos mais utilizados na interacção com uma mesa interactiva [24](Parte 1).	14
2.3	Gestos mais utilizados na interacção com uma mesa interactiva [24] (Parte 2).	15
2.4	Gestos mais utilizados na interacção com uma mesa interactiva [24] (Parte 3).	16
2.5	Área de trabalho privada, feita com uma forma dada à mão. [44].	16
2.6	Exemplos de interacção com uma mesa interactiva [38].	16
2.7	Interacção com um display de grande dimensão [25].	17
2.8	Outro tipo de interacção com um display de grande dimensão [20]	17
2.9	Area Total de trabalho numa mesa interactiva [16].	18
2.10	Viewport Secundário [16].	18
3.1	Processo de desenho para aplicações interativas.	22
3.2	Aplicação de manipulação de fotografias usadas no estudo [26]. (Esquerda) Espaço 3D com fotografias onde se move e altera o seu zoom; (Direita) Rodar a parede de imagens com a capacidade de zoom.	23
3.3	Abordagem FTIR (Frustrated Total Internal Reflection) [2].	25
3.4	Microsoft Kinect [3].	26
3.5	Community Core Vision [1].	26
3.6	TUIO Framework [5].	26
3.7	Representação do procedimento de mudança de escala.	27
3.8	Representação da arquitetura do sistema.	28
3.9	Cópia de um objeto.	30
3.10	Possíveis formas de minimizar	30
3.11	Sincronização de dois objetos.	31
4.1	Botões apresentados aos participantes.	34
4.2	Menu da aplicação depois da melhoria.	39
4.3	Participante a executar a primeira tarefa.	43
4.4	Participante a interagir com a projecção.	43

4.5	Exemplo de um cenário da aplicação.	44
4.6	Ambiente físico da aplicação	45
4.7	(1) Exemplo de menus que a aplicação tem. (2) As cópias feitas pela aplicação	46
4.8	Efeito do número de pessoas na superfície de interação.	48
4.9	Efeito do número de pessoas nas modalidades de interação usadas.	49
4.10	Efeito da abertura da cópia na satisfação das pessoas.	50
4.11	Efeito do local onde aparecem os objetos no uso das superfícies de interação.	51
4.12	Efeito do local onde aparecem os objetos nas modalidades usadas.	52
4.13	Estatística da satisfação dos participantes em relação do aparecimento dos objectos.	53
4.14	Estatística da satisfação dos participantes.	54
4.15	Estatística da satisfação dos participantes acerca do uso da diferentes su- perfícies.	55

Lista de Tabelas

2.1	Diferentes formas dadas à mão e a sua frequência de ocorrência [14]. . . .	9
2.2	Add caption	10
2.3	Gestos identificados para realizar determinadas ações sobre a aplicação [26].	10
2.4	Tempo médio (em segundos) de reflexão para execução de cada ação [26].	12
4.1	Resultados obtidos sobre a representação gráfica dos botões.	36
4.2	Tabela de gestos efetuados na mesa interativa.	37
4.3	Tabela de gestos efetuados na projeção.	38
4.4	Tabela de gestos e comandos de voz dados na mesa interativa.	39
4.5	Tabela de gestos efetuados e comandos de voz dados na projeção.	40
4.6	Explicação dos factores variados nas 6 rondas de teste.	46

Capítulo 1

Introdução

Com o avanço da tecnologia, as formas de interação com os dispositivos e aplicações vão sofrendo alterações. Os telemóveis, tablets e consolas, como a Nintendo Wii ou o Move da PS3, são um exemplo bem vivo de novas formas de interação onde rapidamente o rato e o teclado passaram de moda para dar lugar a uma interação física mais direta. Uma grande inovação no contexto da interação foi o Kinect da Microsoft para a XBox360 onde se pode jogar sem ter nenhum botão ou comando físico, sendo apenas necessário fazer os movimentos com o corpo. Assim, a ideia de interação gestual fica facilitada com o aparecimento do Kinect onde já não são necessários equipamentos intrusivos para rastrear os movimentos das mãos ou do corpo.

Na literatura existe um campo de pesquisa denominado interação natural onde as pessoas interagem com sistemas utilizando uma linguagem corporal que é compreendida pela máquina. Um outro tipo de interação que também tem sido alvo de investigação é o discurso, que tem como objetivo possibilitar interagir com os dispositivos através de comandos de voz. Nos últimos anos houve uma melhoria significativa no reconhecimento da fala pelos dispositivos, o que permite ao utilizador escolher este tipo de interação sem que seja difícil a sua utilização. No seguimento destes estudos pensou-se na junção destes dois tipos de interação, interação por gestos com comandos de voz.

É com base nestes avanços que é proposto este trabalho que tem como objetivo estudar o comportamento dos utilizadores em cenários que juntem interação gestual e interação por comandos de voz em múltiplas superfícies. As superfícies utilizadas irão permitir que se troque informação entre elas. Assim, pretende-se criar um cenário de colaboração entre utilizadores e observar como estes desempenham as suas tarefas.

1.1 Projecto VISTA

O Projeto VISTA (Vision Based Touch Interaction Anywhere) é um projeto nacional que teve início em Janeiro de 2010 e termina em Setembro de 2012. O seu grande objetivo é estudar a interação gestual em várias das suas vertentes, ou seja estudar interação com

e sem toque em múltiplas superfícies e conseguir, também, fazer o tracking das mãos e dedos utilizando equipamentos de baixo custo, como por exemplo webcams. Inicialmente imaginaram-se três cenários de estudo diferentes: (1) projeção visual e interação que partilhem a mesma superfície (exemplo: mesa interativa); (2) projeção visual feita numa superfície e a interação realizada noutra; (3) interação baseada apenas em gestos sem toque, ou seja, sem que haja uma superfície de contacto direto (o utilizador faz gestos que são captados por uma câmara para o tracking das mãos e dedos).

Foram realizados alguns trabalhos de pesquisa no projeto VISTA: “Interação por toque em múltiplas superfícies” por António Regueiras Neto em 2009 [29] que, antecedendo o início oficial do projeto, analisou quais os gestos mais usados quando se interagia com ecrãs de grande dimensão, quais as técnicas para se adaptar a esses ecrãs e qual o impacto da interação gestual em cenários cooperativos; “Estudo de modos de comando em cenários de interação gestual” por Tiago Uttini Gomes em 2010 [15], trabalho que se focou em estudar a diferença entre gestos de desenho e gestos de comando numa aplicação de desenho em superfícies de diferentes dimensões, tais como tablets, mesas interativas e smartboards; “Interação gestual sem superfícies de apoio” por Joana Raimundo Neca em 2011 [27], trabalho que estudou a interação gestual com projeções sem toque e com voz como adição à interação. Por fim será apresentado este trabalho que tem como objetivo estudar a interação gestual em cenários de colaboração com múltiplas superfícies com e sem toque com a ajuda de comandos de voz.

1.2 Motivação

Desde que surgiram computadores para a manipulação e tratamento de dados, que surgiu a necessidade de os controlar de uma forma idêntica à que usamos para manipular objetos reais, por isso temos, por exemplo, o ‘Ambiente de trabalho’ de um computador que se assemelha a uma secretária de trabalho. Mover um ficheiro de uma pasta para outra começou por ser um comando apenas escrito em texto, mas com a utilização de interfaces gráficas em complemento com o rato, o utilizador conseguia agora efetuar movimentos mais próximos da vida real, como mover o ponteiro indicador da posição do rato para cima do objeto (ficheiro), carregar numa tecla do rato que indica que pretende agarrar o objeto e de seguida mover o rato para o local onde o pretende colocar, largando o botão do rato para indicar essa intenção.

Este é um típico movimento que estamos habituados a usar desde há muitos anos, mas com os recentes avanços tecnológicos, que criaram superfícies sensíveis ao nosso toque, poderemos começar a interagir com os dispositivos de uma forma diferente, através de gestos (interação gestual), podendo assim libertar-nos da necessidade de utilizar ratos, teclados ou outros dispositivos que não nos proporcionem esta liberdade de interação.

Apesar dos avanços da tecnologia, a interação gestual ainda é alvo de muito estudo

pois é necessário encontrar um nível de usabilidade idêntico ao que já foi atingido com a utilização dos periféricos habituais. Assim é necessário perceber como é que as pessoas interagem com estes novos dispositivos que oferecem interação gestual, percebendo que tipo de gestos os utilizadores preferem realizar, quais os gestos que são mais adequados a cada ação, quais as ações que podem ser realizadas com este tipo de interação e perceber em que cenários se torna vantajoso a utilização de gestos.

Este trabalho pretende estudar a interação com dispositivos que tenham superfícies de apoio ou contacto, como os telemóveis, tablets ou mesas tácteis e também os dispositivos onde não existe contacto com a sua superfície, como telas na parede onde o input do utilizador será efetuado através de um dispositivo, o Kinect, que deteta os movimentos e gestos do utilizador. Outra forma de input estudada é a utilização de comandos vocais, onde os gestos e a voz se complementam para que determinada ação seja efetuada.

1.3 Objetivos

Neste trabalho pretende-se explorar as capacidades oferecidas pela interação gestual, com adição de voz, e as diferenças existentes, nessa interação, entre uma superfície de toque direto e uma superfície de projeção onde o toque não é permitido. Após uma análise da literatura, verifica-se que a interação gestual em combinação com os comandos de voz e a utilização de múltiplas superfícies de trabalho, é um assunto pouco abordado. Assim, o objetivo inicial desta tese é contribuir para aumentar a caracterização, deste tema, nesta área.

Os objetivos incluem também aplicar e validar uma proposta de processo de desenho de aplicações que permitam interação multimodal e que permitam o uso de várias superfícies. Para tal, foram feitos dois estudos que fazem parte das etapas do deste processo proposto.

O primeiro estudo foca-se na observação dos gestos e comandos de voz efetuados em duas superfícies: a superfície de projeção, onde não é permitido o toque; e a superfície multi-toque, como uma mesa-interativa. O objetivo é encontrar um conjunto de gestos e comandos de voz adequados a determinadas ações que permitam a manipulação de objetos virtuais, nas duas superfícies de trabalho, isto é, para cada superfície encontrar um conjunto de comandos de gestos e voz.

O segundo estudo foca-se na análise do comportamento dos utilizadores em aplicações multimodais, multiutilizador e multisuperfície. Para analisar este comportamento procedeu-se à identificação das variáveis que poderiam ter um impacto significativo na utilização da interação gestual. Uma vez que, no tempo disponível para este trabalho, não se podem estudar todas as variáveis desta interação, foram escolhidas para estudo três variáveis. Assim, o objetivo principal deste estudo foi perceber de que forma estas variáveis influenciam no fluxo de trabalho do utilizador e na interação com a aplicação.

1.4 Contribuições

Deste trabalho resultaram algumas contribuições, nomeadamente:

1. Protótipo de uma aplicação de interação gestual com comandos de voz.
2. Definição de um mapeamento de ações/gestos e de ações/comandos de voz, para serem utilizados nos cenários mais comuns de interação individual sem superfície de apoio, no contexto de uma interação com uma superfície de projeção.
3. Definição de um mapeamento de ações/gestos e de ações/comandos de voz, para serem utilizados nos cenários mais comuns de interação individual em superfície multi-toque.
4. Protótipo de uma aplicação numa mesa interativa com a junção de uma projeção com base no mapeamento de ações/gestos e de ações/comandos de voz anterior.
5. Estudo e caracterização da interação, desenho de aplicações e desenho de interfaces para aplicações multimodais, multiutilizador e multisuperfície.
6. Validação parcial do processo de desenho proposto neste trabalho

O trabalho realizado proporcionou contribuições para a comunidade científica na forma de artigos científicos, nomeadamente:

1. Andreia Ribeiro, Carlos Duarte. (2012) **Move and Speak: Interaction with Multiple Surfaces using Gestures and Voice**. Proceedings of ISPD 2012 - International Symposium on Pervasive Displays 2012, Porto, Portugal, 2012
2. Carlos Duarte, Andreia Ribeiro, Rafael Nunes. (2012) **Designing Multimodal Interactive Systems**. Proceedings of IHCI 2012 - IADIS International Conference Interfaces and Human Computer Interaction 2012, Lisbon, Portugal, 2012

1.5 Estrutura do documento

Este documento está organizado da seguinte forma:

- **Capítulo 2 - Trabalho relacionado**

Neste capítulo é indicado o contexto onde este tema se insere e é feita a síntese de todo o processo de investigação sobre interação gestual, com e sem superfícies de apoio e também com e sem interação por voz. São também indicados quais os gestos mais comuns no processo de interação e quais foram as principais dificuldades encontradas nos vários estudos consultados.

Será também referida a importância da interação entre vários utilizadores, numa mesma superfície ou em superfícies separadas, num contexto de colaboração e comunicação.

- **Capítulo 3 – Desenho de Aplicações Multimodais, Multiutilizador e Multisuperfície:**

Neste capítulo discute-se o desenho de aplicações multimodais e apresenta-se um processo de desenho que aproveita a disponibilidade de novos dispositivos interativos para construção de protótipos e que permitem uma melhor imersão do utilizador no cenário de interação. Descrevem-se ainda os protótipos utilizados nos estudos correspondentes às etapas do processo de desenho apresentado.

- **Capítulo 4 - Estudos de Interação:**

Neste capítulo apresentam-se os resultados das etapas do processo de desenho. A primeira, que se adapta à etapa de compreensão da interação em ambientes sem restrições tecnológicos, corresponde ao primeiro estudo efetuado e a segunda, que se adapta à validação da interação com protótipo de alta-fidelidade corresponde ao segundo estudo.

- **Capítulo 5 - Conclusão e Trabalho Futuro:**

Por último, neste capítulo são apresentadas as conclusões do trabalho e as perspectivas de trabalho futuro.

Capítulo 2

Trabalho relacionado

Este capítulo tem como objetivo apresentar o contexto em que o tema do projeto se insere. É apresentada uma síntese de todo o trabalho de investigação realizado, sobre interação gestual com e sem comandos de voz, interação gestual com e sem superfícies de apoio e também sobre colaboração e comunicação.

2.1 Interacção Gestual

Esta secção mostra as diferenças entre interação gestual com e sem voz e quais os pontos mais importantes a ter em conta além das dificuldades inerentes a esta interação.

2.1.1 Interacção Gestual sem Voz

Nos dias que correm é muito comum encontrarmos aplicações que ofereçam a possibilidade de interação gestual. Aparelhos com interação por toque, tais como smartphones ou mesas interativas, são um bom exemplo disso.

O objetivo é conseguir fazer ações básicas através de gestos que manipulem os objetos virtuais em todos os tipos de superfícies de interação, ou seja, interagir com qualquer dispositivo sem que haja necessidade de utilizar dispositivos de input habituais, como o rato e o teclado.

Nesta secção iremos discutir e identificar quais os gestos mais comuns para as ações básicas e quais as dificuldades encontradas na interação através de gestos.

Gestos mais comuns

A interação gestual foi alvo de estudo nas últimas décadas e hoje em dia consegue-se utilizar quase todas as formas de gestos humanos para interagir de forma natural com computadores na maioria dos domínios da informática.

A interação por gestos tem de ser feita de forma natural e intuitiva, o que significa que o utilizador não necessita de perder tempo a pensar: “Como é que se faz isto?”, quando

quer realizar uma ação sobre determinado objeto virtual. É assim necessário um estudo prévio de quais os gestos a realizar para determinadas ações, e torná-los gestos padrão. Para utilizar os gestos como meio de interação foi necessário estudos que compreendessem os gestos [28, 6] e que melhorassem o seu reconhecimento [45, 23, 43, 41].

Existem propostas de diferentes tipos/estilos de gestos [13, 12, 30]. Uma das propostas [12] distingue os gestos em quatro tipos: deícticos, manipulativos, semaforicos e gesticulados. Numa breve descrição os gestos deícticos, que são utilizados no trabalho “Put-That-There” [7], envolvem apontar para o objeto dentro de um determinado contexto; os gestos manipulativos acontecem quando o gesto efetuado se relaciona com a entidade que se quer manipular; os semaforicos definem-se como um sistema de gestos onde se aplica um dicionário estático ou dinâmico de gestos [30]; e, por fim, os gesticulados são os gestos que utilizamos para combinar com o discurso [13] e tipicamente são utilizados em interfaces de fala e de gestos.

Muitos dos gestos já se encontram padronizados e são muito utilizados em pequenos dispositivos como tablets, smartphones e consolas de jogos, uns já definidos pelo dispositivo, por serem os mais intuitivos e o utilizador os empregar normalmente [42], outros por serem aprendidos mais tarde com o uso do dispositivo. Assim existe uma tendência para usar esses mesmos gestos em qualquer outra aplicação, que tenha as mesmas características.

Apesar do uso destes dispositivos ser comum, existem algumas atividades que se tornam difíceis de realizar através de gestos, ou seja, ainda não é intuitivo para todos os utilizadores qual o gesto mais apropriado para realizar determinado objetivo. Por isso foram realizados estudos que avaliam quais os gestos adequados para cada comando [14, 26, 11].

Estes estudos têm como objetivo observar o comportamento do utilizador em determinados cenários de utilização (interação com e sem toque). É necessário saber qual o gesto utilizado para fazer determinada ação e o que difere de utilizador para utilizador.

Os comandos mais frequentes que são utilizados na interação com um computador são: abrir ficheiros/pastas, arrastar/mover, selecionar texto, rodar, aumentar (zooming), copy/paste, scrolling, undo, redo, etc. [14]. Se o utilizador está habituado a fazer estes e outros comandos com a ajuda do rato e do teclado, então ele esperará conseguir fazer estes mesmos comandos através de gestos e consequentemente fará o gesto que lhe pareça mais conveniente para essa ação.

Julien Epps, Serge Licham e Mike Wu estudaram as formas que o utilizador faz à mão durante a interação com um tabletop numa interação [14]. Este comportamento é importante uma vez que estas formas podem manifestar-se em quase toda a interação gestual com e sem toque [26, 9, 11, 42]. Na imagem 2.1 são apresentadas formas comuns dadas às mãos durante a interação com uma mesa e a sua frequência de utilização.

Mas apesar de ser importante a forma que o utilizador dá à sua mão durante a interação com o dispositivo, é necessário estudar todo o movimento do corpo e verificar os gestos











Hand shape	Example	% of total	Main uses
Index finger		70.1%	Single selection, opening, drawing (>90%), text selection, slider moving (>80%), scrolling (>70%), multiple selection, moving, rotation, zooming, floating menu (>60%), cut/copy (>50%)
Spread hand		20.0%	Rotation (38%), multiple selection (36%), zooming (32%), scrolling (30%), floating menu (28%), drawing (25%)
Flat hand		11.4%	Scrolling (25%), copying (20%), rotation, moving (15%), drawing (13%), zooming (12%), text selection, cut, floating menu (10%)
Grab/Release		4.6%	Cut (25%), copy (13%), moving icons, moving slider (8%)
Vertical hand		1.8%	Cut (8%), text selection, copy (3%)
Fingers together		1.7%	Opening, zooming (5%), text selection, moving, cut, copy, slider moving (3%)
Fist		1.5%	Floating menu (20%), zooming (3%)
'L' shape		1.0%	Floating menu (5%), multiple selection, copy (3%)
'C' shape		0.6%	Floating menu (3%), zooming (2%)
Curved hand		0.6%	Cut (5%), copy (3%)

Tabela 2.1: Diferentes formas dadas à mão e a sua frequência de ocorrência [14].

como um todo (mãos, braços, cabeças e tronco) para que esta interação seja o mais natural possível.

Outro estudo [26], que não se focava apenas na forma das mãos mas no corpo como um todo, mostrou vários utilizadores a interagir em dois cenários diferentes, uma mesa e uma parede com fotografias, onde os utilizadores teriam que fazer diversas ações sobre essas fotografias.

Na Figura 2.1 podemos observar as diferentes formas que foram observadas para realizar o comando de Ampliar.

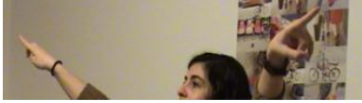


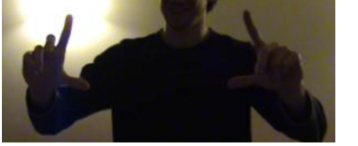
Fingertips		All fingers apart	
Fingers together		Fingers in a triangular shape	

Figura 2.1: Diferentes formas de ampliar [26].

A Tabela 2.3 apresenta, como resultado, os gestos realizados pelo utilizador para cada comando/ação pretendida. Algumas ações não têm gestos padrão e existem ações que têm gestos repetidos, logo é necessário ser feita uma desambiguação destes gestos, que é

conseguida através da adição de comandos de voz, como será mostrado mais à frente.

Tabela 2.2: Add caption

Ações	Descrição do Gesto
Zooming	(zoom in) Dois pontos paralelos à superfície afastando-se um do outro. (zoom out) Dois pontos paralelos juntando-se um ao outro.
Sobrepôr	Um ponto movendo-se para longe da superfície.
Mover	Um ponto movendo-se paralelamente à projeção.
Parar de mover	Um ponto imóvel.
Rodar	Um ponto centrado no objeto e outro ponto descrevendo o movimento de rotação iniciando no primeiro ponto. Um ponto apenas, descrevendo o movimento de rotação.
Apagar	Desenhar um X. Bater palmas uma vez. Fazendo um gesto de rasurar. (movendo um ponto para a esquerda e para a direita).
Desfazer	Um ponto da esquerda para a direita.
Refazer	Um ponto da direita para a esquerda.
Anterior	Um ponto da direita para a esquerda.
Seguinte	Um ponto da esquerda para a direita.

Tabela 2.3: Gestos identificados para realizar determinadas ações sobre a aplicação [26].

Neste estudo observou-se que para efetuar uma ação, como ampliar, os utilizadores

recorreram tendencialmente ao mesmo gesto, mas alguns usaram, por exemplo, as mãos abertas e outros as mãos fechadas, como é mostrado na Figura 2.1. Assim, é necessário ignorar a forma dada às mãos, pelo menos para a maioria dos gestos, identificando-as apenas como pontos, e retirar a relevância dada à forma como o utilizador coloca a mão. Na próxima secção serão discutidos os problemas da interação gestual.

Dificuldades encontradas

Uma das grandes dificuldades encontradas na interação gestual é o facto de os utilizadores fazerem gestos idênticos para diferentes ações [26]. Isto torna bastante difícil para a aplicação saber qual é a ação que o utilizador deseja efetuar, nem sempre sendo o contexto suficiente para suportar essa decisão. Uma solução para resolver este problema seria a aprendizagem do utilizador, ou seja a aplicação tinha os seus gestos padrão e utilizadores inexperientes aprendiam quais os gestos que teriam que usar. Mas esta solução não está de acordo com o objetivo deste trabalho, pois pretende-se oferecer uma interação intuitiva e natural e não obrigar o utilizador a aprender os gestos.

Outra questão que se coloca na interação por gestos sem superfícies de apoio é saber quando os gestos são dirigidos ou não à aplicação. Isto é, existe necessidade de se saber se o utilizador quer realizar um gesto que seja um comando de interação ou simplesmente se está a fazer um gesto porque está a ter uma conversa com outra pessoa.

Para resolver estes problemas pensou-se em adicionar comandos de voz. Na próxima secção deste capítulo falaremos das vantagens e desvantagens desta adição à interação por gestos.

2.1.2 Interação Gestual com Voz

A interação gestual por si só tem alguns problemas, como foi abordado na secção anterior. Para ultrapassar estes problemas pensou-se na junção de comandos de voz [26, 38, 8, 31]. Esta interação por voz irá facilitar a interação gestual uma vez que é um complemento para esta e permite desambiguar alguns gestos, ou seja, conseguir que o utilizador não sinta dificuldade em realizar um determinado gesto.

Com a adição dos comandos de voz foi observado que o propósito dos gestos seria modificado [26, 7, 38]. Com esta fusão os gestos serviam apenas para complementar os comandos de voz. Por exemplo: para apagar algo o utilizador apenas diria 'DELETE' e apontava para o objeto que queria apagar.

Na secção seguinte será identificado que tipos de gestos se modificam com a junção de comandos de voz, quais os que não sofrem qualquer alteração e quais as vantagens desta integração.

Vantagens da integração de voz

Pela análise da tabela 2.3 pode observar-se o que os utilizadores efetuam gestos diferentes para a mesma ação. Com a adição de comandos de voz isso muda pois o gesto servirá para transmitir a ação a ser executada ou apenas apontar para o alvo da ação [38]. As ações que provocam dúvida aos utilizadores de qual o gesto a utilizar serão efetuadas através de um comando de voz. Exemplos dessas ações são: apagar, desfazer, refazer e rodar, que como observado não teriam um gesto padrão [26].

O “Put-that-there” [7] é outro estudo que mostra a integração do discurso na interação e chegou à conclusão que os comandos de voz geram a ação e os gestos apontam para o sujeito da ação.

A principal vantagem desta combinação de gesto e voz é conseguir uma interação mais natural e fluída [38, 8, 26, 7], pois nós humanos quando interagimos com uma pessoa fazemos uso da linguagem gestual e da fala. Assim se alcançarmos esta interação com máquinas, melhor e mais intuitiva esta o será.

Ação	Apenas Gestos	Gestos e Voz
Zoom in	1,3	0,8
Zoom out	0,2	0,2
Mover	1,3	1,1
Parar de mover	0,5	0,4
Sobrepôr	3	3,1
Rodar	1,1	1,4
Apagar	7,1	0,4
Desfazer	7	3,2
Refazer	0,6	1

Tabela 2.4: Tempo médio (em segundos) de reflexão para execução de cada ação [26].

Na Tabela 2.4 são comparados os tempos de reflexão¹ utilizando apenas gestos e utilizando gestos e voz. A conclusão que se tira desta informação é que na interação gestual com comandos de voz os utilizadores executam o comando mais rapidamente, o que também é indicativo que a interação se torna mais intuitiva.

¹Tempo que os participantes no estudo demoraram a pensar sobre como executar o comando com as modalidades que tinham disponíveis.

Problemas inerentes

Como nem tudo é perfeito, a junção de comandos de voz pode provocar alguns dissabores no que toca ao reconhecimento de discurso. Este não é um problema muito grave pois este assunto tem sido alvo de pesquisa nos últimos anos [18, 32, 21] e, hoje em dia, já se faz um reconhecimento de discurso bastante eficiente, mas em alguns casos pode gerar problemas devido às características vocais do utilizador. Se o utilizador tiver um timbre de voz mais baixo, ou um sotaque diferente do habitual [21] por vezes os comandos de voz podem ser mal interpretados ou mal “ouvidos” pela aplicação e isso provocará alguma frustração no utilizador.

Outro problema consiste no mesmo que acontece com os gestos: separar discurso entre humanos e interação entre máquina-humano. É necessário abordar uma estratégia para que a máquina saiba quando o discurso é dirigido para ela, como por exemplo dar comandos de iniciação e fim de discurso, ou seja comandos como: “Start Listening” e “Stop Listening” como é o caso do reconhecedor de comandos de voz do Windows [22]. O ideal seria identificar automaticamente quando o utilizador está a interagir com a máquina e quando não está, mas por enquanto este cenário não é possível. Numa outra área de investigação chegou-se à conclusão que os humanos têm a tendência de dar nomes aos robots e máquinas que têm em casa [35], assim podemos dizer que uma boa estratégia para distinguir as várias superfícies seria dar nomes a cada uma delas.

2.2 Tipos de Interação

Os tipos de interação aqui estudados referem-se às superfícies com que o utilizador está a lidar. Uma superfície de contacto direto tem um tipo de interação diferente de uma superfície que não está ao alcance físico do utilizador.

Nesta secção serão descritos estes dois tipos de interação e serão também mostradas as principais diferenças com exemplos práticos. Será discutida a integração de voz nas duas superfícies, como um complemento para o controlo dos dispositivos, dando também ênfase ao feedback fornecido.

2.2.1 Interação com uma superfície de apoio

As superfícies de apoio são uma das formas de interação tangível, que pertencem às TUI's (Tangible User Interface). Interação tangível é um conceito pertencente à área da Interação Pessoa-Máquina (HCI – Human-Computer Interaction) que envolve estudo, planeamento e design da interação entre pessoas e computadores. Um dos exemplos de interação com uma superfície de apoio é a interação com uma mesa interativa. Foi necessário haver estudos desta interação para perceber como os utilizadores utilizavam este novo meio de interagir e se se sentiam à vontade em utilizá-lo [19, 40, 33, 10, 44, 17].

Destes estudos resultaram questões de design a ter em conta, por exemplo, como suportar múltiplas orientações numa determinada área [40].

Na interação com uma superfície de contacto direto, o utilizador interage com o dispositivo/mesa utilizando os dedos ou outro tipo de apontador, fazendo gestos que se reproduzem em ações como, por exemplo, arrastar ou seleccionar, existindo feedback no local onde o dispositivo foi “tocado”, para que o utilizador tenha a noção que a operação efetuada, corresponde à sua intenção.

Meredith Ringel Morris, Jacob O. Wobbrock e Andrew D. Wilson [24] tentaram perceber quais os gestos que os utilizadores mais realizavam para cada ação. O resultado obtido é mostrado nas figuras 2.2, 2.3 e 2.4

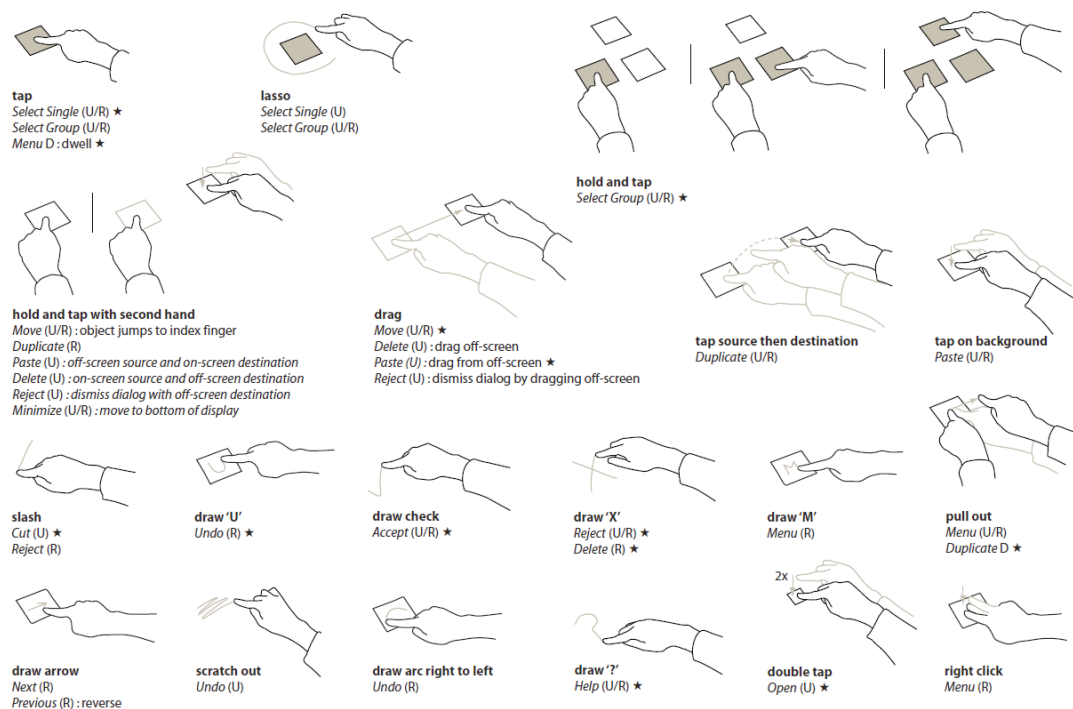


Figura 2.2: Gestos mais utilizados na interacção com uma mesa interactiva [24](Parte 1).

No trabalho de Mike Wu e Ravin Balakrishnan [44] é apresentada uma aplicação “RoomPlanner” que incorpora várias técnicas de interação colaborativa e explora o uso de multi-toque, formas dadas às mãos e input de gestos. Este estudo analisou as várias técnicas de interação usadas pelo utilizador que são: utilização de apenas um dedo; utilização de dois dedos; utilização da forma de uma mão; e utilização da forma dada às duas mãos.

Com esta análise foi observado que os utilizadores para além de trabalho colaborativo também mostraram competitividade em fazer as atividades construindo uma área de trabalho privada (Figura 2.5). Assim podemos perceber que ter uma área de trabalho privada é uma ideia que não descontenta o utilizador.

Num outro estudo [33], onde o objetivo era fazer com que os alunos adquirissem

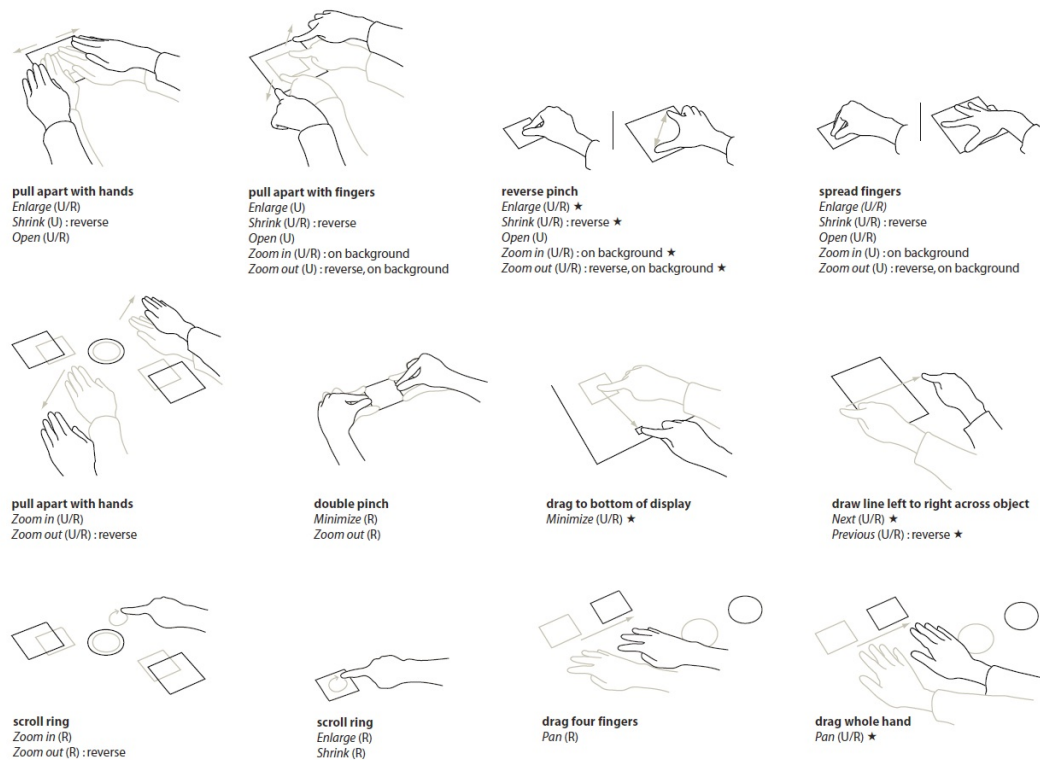


Figura 2.3: Gestos mais utilizados na interacção com uma mesa interactiva [24] (Parte 2).

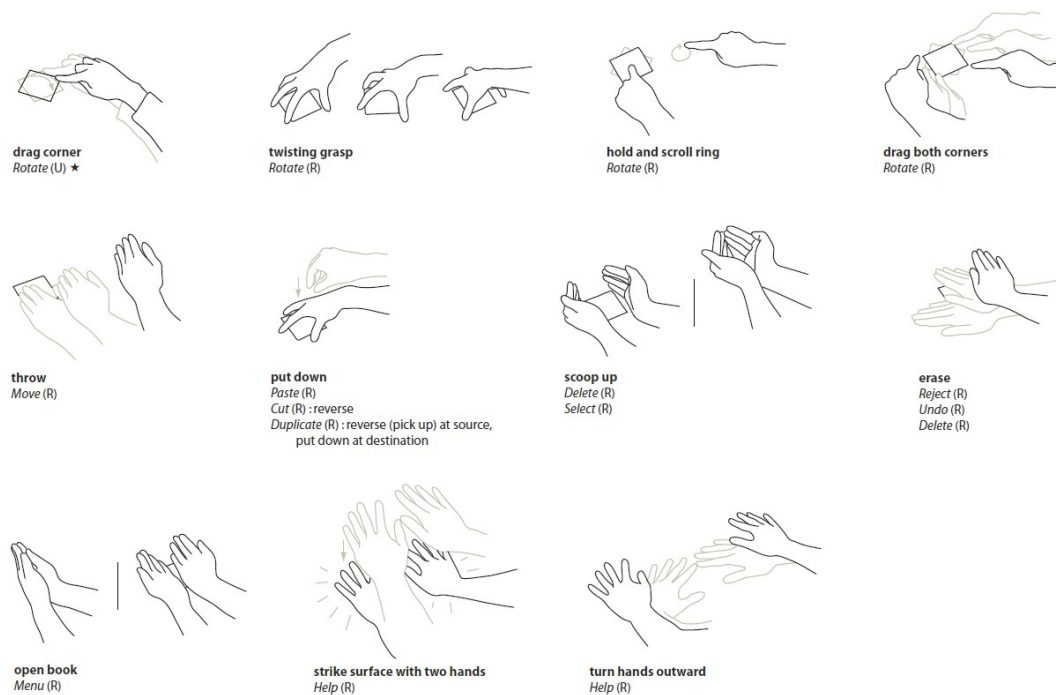


Figura 2.4: Gestos mais utilizados na interacção com uma mesa interactiva [24] (Parte 3).

conhecimento sobre uma área de Biologia, os genomas, observou-se que a utilização da mesa interactiva melhora o desempenho dos alunos, reduz a carga de trabalho e aumenta a



Figura 2.5: Área de trabalho privada, feita com uma forma dada à mão. [44].

diversão. Assim podemos concluir que a utilização deste tipo de superfícies pode trazer muitas vantagens quando utilizada no trabalho ou na educação e não apenas no lazer.

A utilização de voz neste tipo de interação tem o mesmo objetivo que vimos na secção anterior. Através do trabalho de Edward Tse, Saul Greenberg e Chia Shen [38, 39] conseguimos observar que o discurso é usado com os gestos/toque para fornecer comandos específicos do que se quer fazer. Na figura 2.6 podemos ver exemplos de interação com a mesa utilizando comandos de voz.

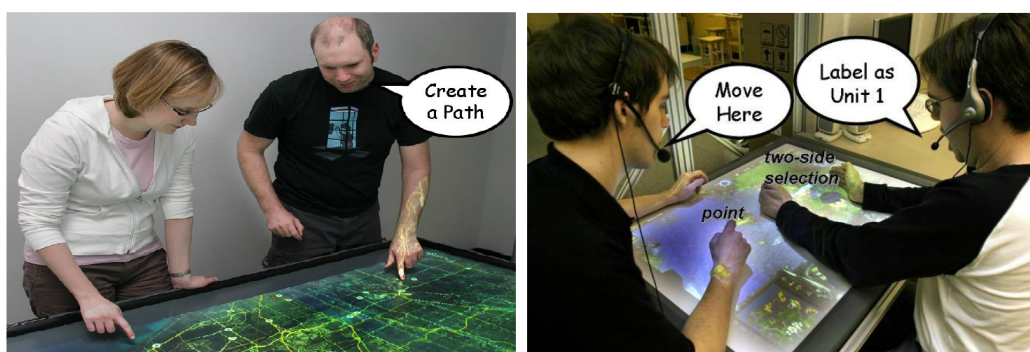


Figura 2.6: Exemplos de interacção com uma mesa interactiva [38].

2.2.2 Interação sem uma superfície de apoio

A interação sem uma superfície de apoio implica que o utilizador não tenha contacto físico com a superfície, ou seja, a superfície encontra-se distante do utilizador. Desta forma o input será feito através de gestos e voz. As superfícies que se encontram a uma certa distância do utilizador normalmente são ecrãs ou projeções com um tamanho acima do que é habitual.

A interação com este tipo de superfícies, como é de esperar, é um pouco diferente do habitual pois a área de interação é muito maior.

Alguns estudos [20, 36, 25] analisam a facilidade ou não do uso destas superfícies.

Uma das estratégias abordada por Christian Müller-Tomfelde, Kelvin Cheng e Jane Li [25] é o uso de um toque “pseudo-directo” que permite interagir, com a superfície à distância, através de um quadro de toque transparente (Figura 2.7). Os pontos de contacto no quadro são projetados sobre a superfície, para que os utilizadores tenham a impressão de tocar na tela grande diretamente. Neste trabalho não é isto que se pretende mas o

utilizador também tem um limite de movimentos pois os seus braços não são do tamanho da tela projetada e tem de existir uma adaptação de escala à projeção.

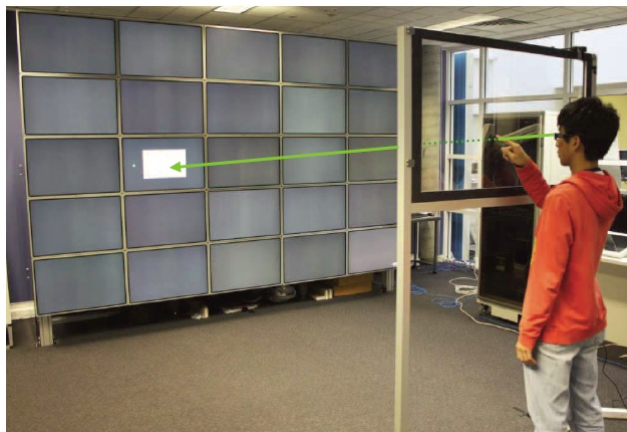


Figura 2.7: Interação com um display de grande dimensão [25].

Uma outra abordagem com ecrãs de grande dimensão sem toque é a de Shahzad Malik, Abhishek Ranjan e Ravin Balakrishnan [20] onde a interação é feita numa superfície perto do utilizador e a projeção do que é realizado é feita num ecrã de grande escala longe do utilizador (Figura 2.8). Os autores dizem que o tipo de interação é semelhante ao uso de uma mesa interativa, pois existe toque nas ações do utilizador. Esta é uma abordagem que não se enquadra neste trabalho mas que pode ser útil para perceber como é usado um ecrã de grande dimensão pelos utilizadores. Neste caso existe uma pequena área no ecrã que é um “zoom in” da área onde o utilizador quer efetuar ações.

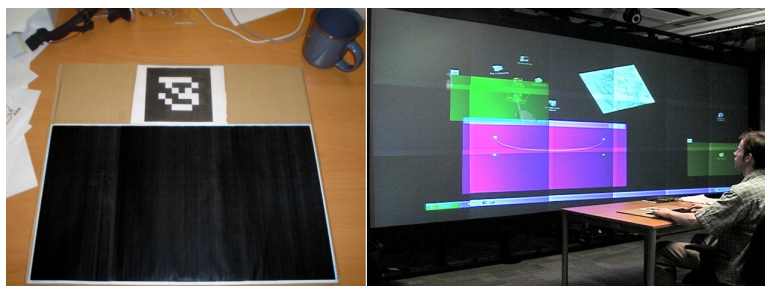


Figura 2.8: Outro tipo de interação com um display de grande dimensão [20]

O estudo de Desney S. Tan, Darren Gergle, Peter G. Scupelli e Randy Pausch [36] prova que o uso de ecrãs de grandes dimensões pode melhorar o desempenho espacial do utilizador, assim ter um destes ecrãs é uma mais valia para este trabalho.

Neste trabalho pretende-se que a interação com a superfície que se encontra à distância seja através de gestos e voz como acontece no trabalho de Joana Neca e Carlos Duarte [26].

As ações pretendidas pelo utilizador terão de ser reconhecidas pelo dispositivo através de gestos, mas dependendo da distância a que o utilizador se encontra, estas poderão

tornar-se ambíguas. É então necessário que se complementem os gestos com comandos de voz, que reforçam o que se pretende dizer [26].

Um problema que daqui poderá advir é o facto de existir um vocabulário extenso e não haver para já um consenso sobre o que o utilizador terá de dizer para efetuar uma ação. Várias palavras poderão designar a mesma ação, logo terá de existir uma boa base de vocabulário, que tenha “inteligência” suficiente para saber detetar o contexto da ação pretendida.

2.3 Colaboração/Comunicação

Os cenários de colaboração podem ser facilitados com a introdução do tipo de interação abordado neste trabalho: a interação por gestos e voz em múltiplas superfícies. O design de superfícies de interação para cenários de colaboração é difícil uma vez que não dá para prever como os grupos de trabalho coordenam as suas ações quando estão perante uma mesa interativa [37].

Carl Gutwin e Saul Greenberg [16] mostram no seu trabalho uma aplicação onde os utilizadores dividem a área total de trabalho (Figura 2.9), através de viewports secundários (Figura 2.10), e cada um tem a sua função. O resultado deste trabalho mostra que esta abordagem com viewports secundários é uma mais-valia pois os utilizadores conseguiram interagir sem grandes problemas com a aplicação.

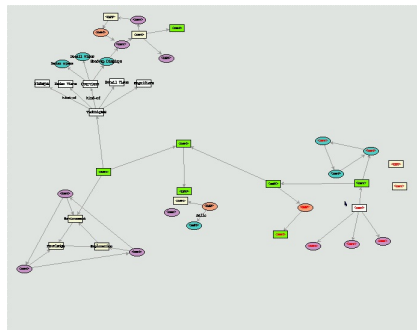


Figura 2.9: Area Total de trabalho numa mesa interactiva [16].

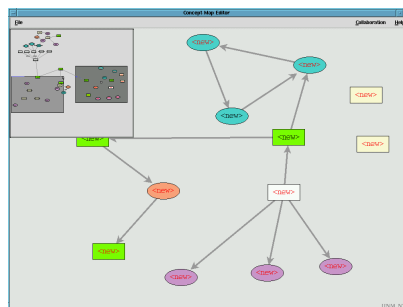


Figura 2.10: Viewport Secundário [16].

Através dos estudos efetuados por [38, 40, 33, 10, 44, 17, 39, 25] podemos perceber que o tipo de interação de gestos e voz com a utilização de superfícies interativas pode melhorar bastante o comportamento colaborativo dos utilizadores podendo até haver um aumento de produtividade em trabalhos realizados em grupo.

A integração de comandos de voz pode alterar um pouco a forma como os utilizadores se comportam nestes cenários, pois se o dispositivo vai estar a “ouvir” então a comunicação entre utilizadores poderá estar condicionada.

2.4 Discussão

A pesquisa efetuada mostrou que é importante a junção de gestos com voz na interação com superfícies que não estão ao alcance físico do utilizador. Os gestos efetuados sem a adição de voz podem tornar-se um pouco ambíguos e pouco naturais pois cada ação pode ter associada mais de um gesto. Com a adição de comandos de voz, este problema já não acontece e a interação torna-se mais natural e intuitiva. As diferenças de interação entre as diferentes superfícies também são um ponto importante a focar, pois existem diferenças que se devem ter em conta quando se quer construir uma aplicação com superfícies distintas.

A interação com superfícies de apoio requer o toque, como foi visto nos vários trabalhos apresentados. A interação sem superfícies de apoio ainda não está bem definida. Neste trabalho iremos estudar a interação com estas superfícies através de gestos e voz realizados à distância. Na próxima secção será abordada a problemática do desenho de aplicações em cenários interativos com estas características, levando aos estudos realizados no âmbito deste trabalho.

Capítulo 3

Desenho de Aplicações Multimodais, Multiutilizador e Multisuperfície

Neste capítulo discute-se o desenho de aplicações multimodais e apresenta-se um processo de desenho que aproveita a disponibilidade de novos dispositivos interativos para construção de protótipos e que permitem uma melhor imersão do utilizador no cenário de interação. Com base neste processo de desenho é desenvolvida uma aplicação que irá permitir o estudo de algumas variáveis a ter em conta neste tipo de aplicações, uma vez que não se poderão estudar todas num só trabalho. Essa aplicação é descrita na última parte do capítulo.

3.1 Processo de Desenho

As novas tecnologias de interação estão a tornar-se pervasivas hoje em dia. O reconhecimento de gestos está disponível não só em tablets e smartphones, mas também em sistemas de jogos como o Kinect, da Microsoft. O reconhecimento de voz também se está a tornar dominante em smartphones mais recentes. Na área do entretenimento tem existido maior trabalho de desenvolvimento desses sistemas [34]. No entanto, estas tecnologias estão a chegar a um ponto de amadurecimento que os torna disponíveis e desejáveis para outros campos [17, 24, 40]. O objectivo deste trabalho é perceber como as pessoas se vão adaptar a estas tecnologias de interação noutros campos, como substitutos para os dispositivos de entrada tradicionais. Para cumprir este objetivo, começamos por observar como as pessoas usam esta tecnologia sem limitações técnicas impostas como parte de um processo de desenho.

O processo de desenho aqui abordado é um processo iterativo que apresenta três etapas: (1) identificar as funcionalidades a disponibilizar na aplicação e desenhar os mecanismos de interação necessários para concretizar essas funcionalidades; (2) desenhar e implementar um protótipo de alta fidelidade que permita a utilizadores interagirem com as funcionalidades da aplicação. É fundamental um protótipo de alta-fidelidade para a imersão dos utilizadores na experiência interativa e no cenário que se está a tentar

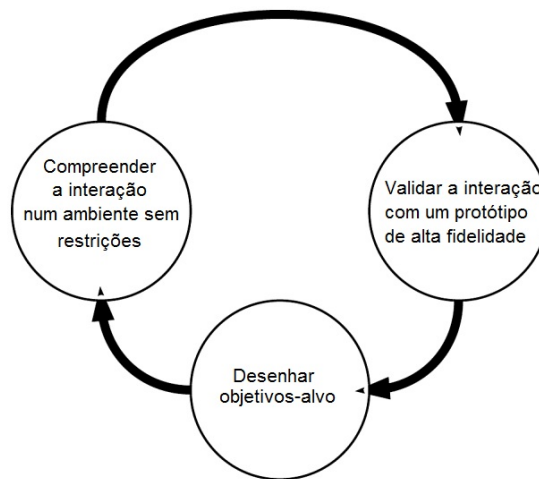


Figura 3.1: Processo de desenho para aplicações interativas.

alcançar. Hoje em dia a tecnologia permite-nos construir protótipos de alta-fidelidade de uma forma mais rápida e com poucos recursos, não havendo assim necessidade de perder tempo com protótipos de baixa fidelidade. O uso da abordagem do Feiticeiro de Oz é imprescindível nesta etapa porque permite entender o comportamento natural do utilizador ao conseguir obter os gestos e comandos de voz, que de outra forma, não seriam facilmente prototipados; (3) construir uma aplicação interativa, reutilizando na medida do possível o protótipo de alta-fidelidade, incorporando os resultados aprendidos na etapa anterior. Isto proporciona uma aplicação que atende aos objetivos da interação, ao sustentar o comportamento natural do utilizador. Este processo de desenho pode ser representado pela figura 3.1.

3.2 Aplicação do Processo de Desenho em Aplicações Multimodais e Multisuperfície

A aplicação deste processo de desenho em aplicações multimodais e multisuperfície tem como objetivo o estudo da interação gestual, com e sem toque, suportada por comandos de voz. Neste momento já foram realizadas duas iterações deste processo. A primeira iteração foi feita num trabalho de mestrado anterior e a segunda iteração está descrita neste trabalho. As secções seguintes mostram o trabalho efetuado na primeira iteração e os objetivos a conseguir com o desenvolvimento desta segunda iteração.

3.2.1 Primeira Iteração do Processo de Desenho

Seguindo o processo de desenho apresentado acima, para o objetivo de entender melhor como as pessoas interagem com várias modalidades gestuais e de voz, é necessário, em primeiro lugar, observar como as pessoas interagem naturalmente com a aplicação, isto é,

sem que haja restrições impostas pela tecnologia. Para isso, foi desenvolvido um estudo [26] em que os participantes foram convidados a executar tarefas em duas aplicações onde era permitido ver fotografias e manipulá-las (ver Figura 3.2). A interação utilizada neste estudo foi a interação gestual, sem ajuda de toque, com adição dos comandos de voz, numa superfície de projeção.



Figura 3.2: Aplicação de manipulação de fotografias usadas no estudo [26]. (Esquerda) Espaço 3D com fotografias onde se move e altera o seu zoom; (Direita) Rodar a parede de imagens com a capacidade de zoom.

No estudo foi utilizada a técnica de Feiticeiro de Oz, onde as aplicações eram controladas em conjunto pelo participante e por um assistente. O participante controlava um cursor apontando para a superfície. O reconhecimento do gesto de apontar foi efetuado através do Kinect da Microsoft. Os gestos emitidos pelo participante foram interpretados pelo assistente, que foi responsável por introduzi-los usando atalhos com o teclado. Usaram-se duas aplicações para entender o impacto que o espaço 3D causava nas tarefas executadas. Uma aplicação era 2D onde as fotos não suportavam sobreposição, apenas poderiam ser vistas através de zoom in e zoom out. A outra aplicação foi efetuada num espaço 3D onde as fotografias poderiam ser sobrepostas e ser movidas livremente nas três dimensões.

Compreender o Uso dos Gestos e Voz

O estudo mostra que as ações que não impliquem a manipulação direta de objetos (como por exemplo, apagar, desfazer ou refazer) requerem o uso de comandos de voz em combinação com o gesto de apontar. Caso contrário os participantes mostram grandes dificuldades na conclusão da ação, e os gestos efetuados eram muito diversificados para serem eficientemente reconhecidos por um sistema de reconhecimento de gestos. Ao usar a voz (combinada com o gesto de apontar) as dificuldades não eram sentidas e os comandos foram comuns à maioria dos participantes.

As ações que manipulavam os objetos diretamente, como mover ou rodar, foram realizadas, principalmente, através de gestos. Foi observado que as pessoas desempenham o mesmo gesto com diferentes significados. Quando isto aconteceu, a introdução de comandos de voz permitiu eliminar essas ambiguidades.

Validar os Resultados

Após o primeiro estudo, foi efetuado em segundo estudo com um apoio tecnológico necessário. Foram utilizados reconhecedores de gestos e de voz, que identificaram todos os gestos e comandos de voz obtidos na etapa anterior do processo. O estudo foi efetuado com um protótipo diferente do primeiro estudo, este consistia em navegar na aplicação Google Earth através de gestos e comandos de voz. A grande diferença foi não ser necessário um assistente que simulasse as ações com atalhos do teclado. A conclusão que se obteve com esta segunda etapa foi a de que os resultados obtidos num ambiente sem restrições se adequam a um ambiente tecnicamente limitado. Os participantes interagiram e realizaram as tarefas sem dificuldades. Os reconhecedores de gestos e comandos de voz foram capazes de interpretar os comandos de forma correta.

3.2.2 Segunda Iteração do Processo de Desenho

A segunda iteração do processo tem como objetivo estudar a interação gestual, com adição de voz, e as diferenças de interação, entre uma superfície de toque direto e uma superfície de projeção onde o toque não é permitido. Na primeira etapa do processo identificam-se as funcionalidades que querem ser analisadas com a interação proposta. As funcionalidades que se pensou para o estudo desta interação foram: manipular objetos diretamente, como mover, rodar e ampliar e utilizar ações abstratas sobre os objetos tais como copiar, minimizar, maximizar, sincronizar, fechar/apagar e enviar para a superfície contrária (troca de objetos entre superfícies).

Na segunda etapa fez-se um procedimento semelhante ao trabalho apresentado na secção anterior, utilizando-se a técnica do Feiticeiro de Oz. Eram dadas tarefas aos participantes e este tinham liberdade de expressão para gestos e comandos de voz, pois não existiam restrições tecnológicas. Em cada teste um assistente simulava cada ação do participante. Este estudo permite obter um conjunto de gestos e comandos de voz para serem aplicados no protótipo de alta-fidelidade da etapa seguinte.

Para a terceira etapa do processo de desenho produziu-se um protótipo de alta-fidelidade que já apresentava, aos participantes, restrições tecnológicas, de acordo com os reconhecedores de gestos e de voz que suportavam o conjunto de comandos adquiridos da etapa anterior.

Assim este trabalho está contextualizado dentro do processo de desenho proposto, pois respeita todas as etapas do processo mostradas na figura 3.1.

No capítulo 4 são apresentados, em mais detalhe, os dois estudos efetuados nesta segunda iteração do processo de desenho.

3.3 Componentes (Hardware)

Para desenvolver os dois protótipos atrás referidos, foi preciso um conjunto de componentes de hardware: uma mesa interativa, para captura de gestos efectuados através de toque, um dispositivo de captura de gestos sem toque e um dispositivo de captura de voz. Para estes dois últimos foi utilizado o Kinect da Microsoft.

A mesa interativa foi construída por uma equipa de alunos do Departamento de Informática, na qual eu estou integrada e funciona utilizando a técnica FTIR (Frustrated Total Internal Reflection). Esta técnica consiste em ter uma superfície em acrílico e em torno desta é colocada uma fita de LED's infravermelhos. A luz infravermelha é refletida através do interior do acrílico (Figura 3.3). Em cima da superfície coloca-se uma folha de papel vegetal ou papel velino, coberta por silicone líquido para uma melhor aderência e para melhorar a transmissão de luz quando algo é colocado sobre a superfície. Quando colocamos o dedo em cima da superfície é enviada luz infravermelha para baixo que é capturada por uma câmara de infravermelhos. Neste caso foi usada a câmara da Playstation 3, a PlayStation Eye, após a remoção do seu filtro de infravermelhos.

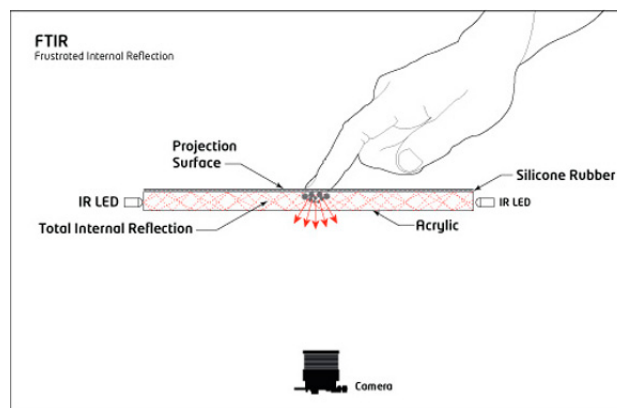


Figura 3.3: Abordagem FTIR (Frustrated Total Internal Reflection) [2].

Para a captura de gestos sem toque e para a captura dos comandos de voz foi utilizado o Microsoft Kinect. O Kinect tem uma câmara RGB, um sensor de profundidade e um emissor de infravermelhos. Como se pode observar na Figura 3.4 também contém quatro microfones, o que permite gravar e encontrar a localização do som. Assim, os gestos e os comandos de voz são recolhidos pelo Kinect.

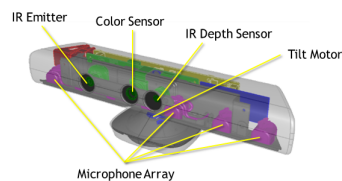


Figura 3.4: Microsoft Kinect [3].

3.4 Tecnologias

Para o desenvolvimento dos protótipos foi necessário recorrer a diferentes tecnologias. Na mesa interativa foi utilizado um programa para fazer o “tracking” e capturar as coordenadas dos “blobs” feitos pelos dedos em contacto com a superfície, o Community Core Vision (CCV)(ver Figura 3.5). O CCV envia os dados recolhidos em formato de eventos TUIO.

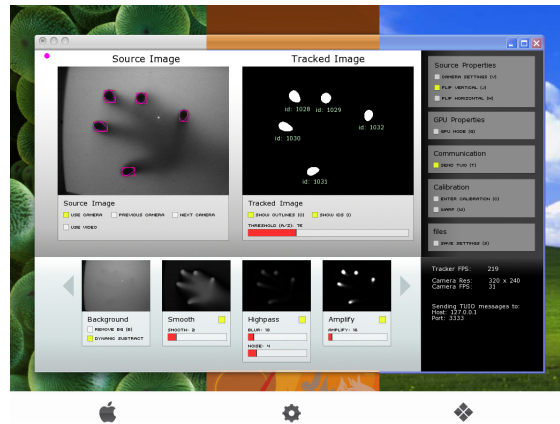


Figura 3.5: Community Core Vision [1].

TUIO é uma framework que define um protocolo para superfícies multi-toque [5]. Os eventos que são capturados pela câmara são enviados para uma aplicação, o “TUIO tracker”, que através do protocolo envia os dados a uma aplicação cliente de TUIO (como podemos observar na figura 3.6). Esta aplicação cliente tem que ser instalada no browser utilizado, neste caso o Google Chrome e para isso foi necessário proceder à instalação de um plugin, o “npTuioClient”.

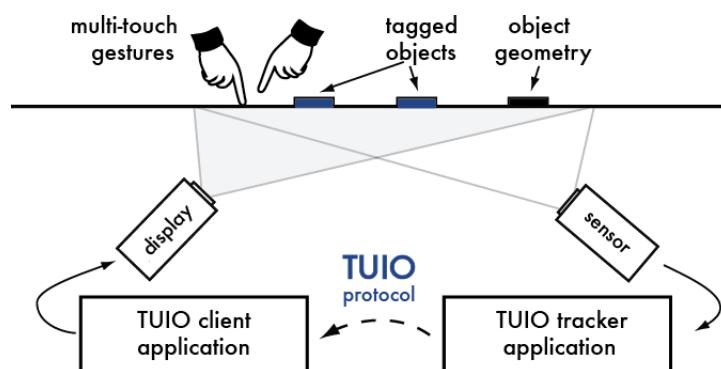


Figura 3.6: TUIO Framework [5].

Para a superfície de projeção foi utilizado o Kinect SDK oficial da Microsoft e foi necessário utilizar uma aplicação em C# (Visual Studio 2010) para reconhecer os gestos. O reconhecedor de gestos utilizado foi o 'Kinect SDK Dynamic Time Warping (DTW)

Gesture Recognition’ disponibilizado no site CodePlex (Project Hosting for Open Source Software)¹.

Para identificar os comandos de voz dados pelas pessoas era necessário ter um reconhecedor. Ter dois reconhecedores diferentes não era uma solução possível, pois existia apenas um Kinect. Então construiu-se uma aplicação em C# que é capaz de fazer os dois tipos de reconhecimento, de gestos e voz.

Decidiu-se construir uma aplicação Web, como protótipo, porque torna-se mais fácil de protótipar em Web do que em aplicações nativas. Assim, não é necessário a instalação de componentes de uma linguagem, nem escolher um sistema operativo específico.

Como o protótipo é uma aplicação Web foi necessário enviar os dados recolhidos pelo reconhecedor, utilizando WebSockets, e escalá-los para serem utilizados no browser. Para este propósito foi utilizado o servidor NodeJs [4], que é um servidor de JavaScript. Foi escolhido porque permite passar um grande volume de dados em tempo real. É também através do servidor NodeJs que os dados da mesa interativa e da projecção são trocados entre si.

Os dados necessitavam de sofrer uma mudança de escala porque os valores enviados, pela aplicação C#, vinham em metros (distância das mãos ao kinect em metros reais) e o browser apenas aceita valores em pixéis. O servidor analisava os dados recolhidos pela aplicação C# e construía uma caixa de limites através do comprimento dos braços do utilizador. Desta forma era conseguida uma equivalência de valores e procedia-se à adaptação desses valores recebidos. A figura 3.7 ilustra o procedimento desta mudança de escala.

3.5 Arquitetura

Para produzir a aplicação organizaram-se os componentes e tecnologias dos pontos referidos acima, da forma que a imagem 3.8 ilustra.

Como pode ser observado os utilizadores interagem com a mesa interativa através de gestos, com toque e por voz. Também podem interagir com a superfície de projecção com gestos sem toque e por comandos de voz. Para captura dos gestos sem toque, é utilizado o Kinect que recolhe os dados. Os dados recolhidos são os comandos de voz ditos pelo utilizador, as coordenadas das duas mãos dos utilizadores e o gesto que foi reconhecido.

Estes dados são enviados com a ajuda de websockets para o servidor nodejs. Este servidor trata de fazer a separação dos dados e transmiti-los às aplicações respectivas. Se o servidor receber gestos então estes são enviados à aplicação sem toque para a representação na projecção. Se os dados recebidos pelo servidor forem comandos de voz o servidor envia essa informação às duas aplicações e cada uma é responsável por empregar esse comando.

Existe um problema com os comandos de voz: para qual das superfícies é o comando

¹<http://www.codeplex.com/>



Figura 3.7: Representação do procedimento de mudança de escala.

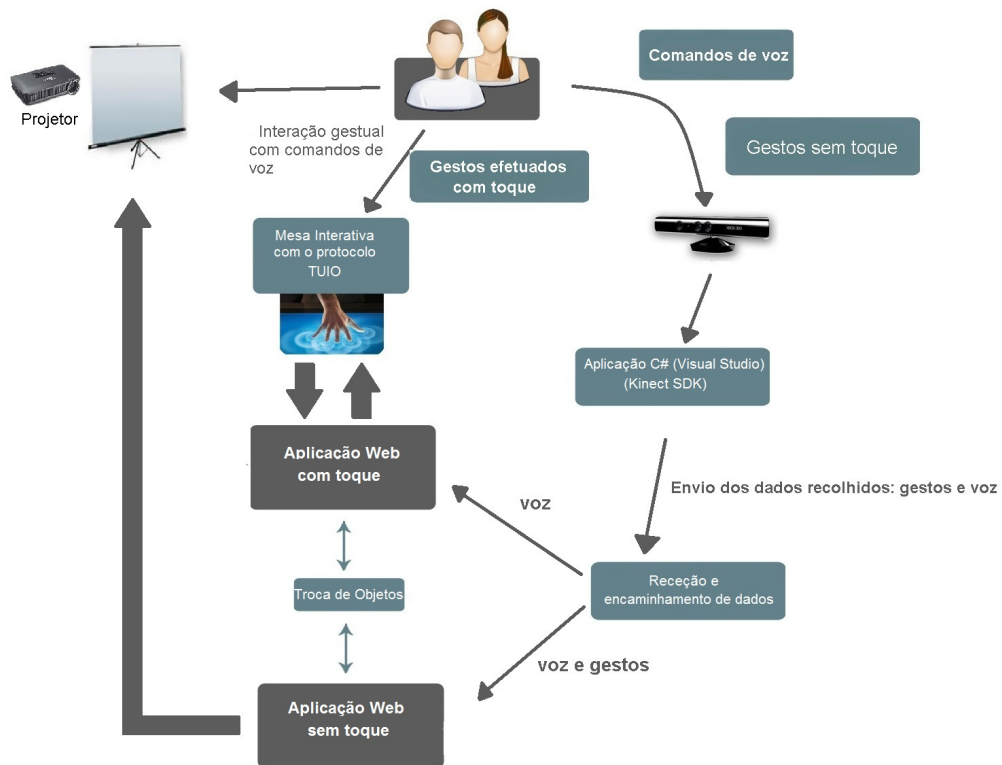


Figura 3.8: Representação da arquitetura do sistema.

que foi dito? Para resolver este problema decidiu-se dar nomes às superfícies. Assim quando um comando era dirigido à mesa interativa este tinha de ser começado com o nome “table”. Se fosse dirigido à projeção então o comando tinha de ser precedido por “kinect”. Por exemplo, para copiar um objeto na mesa o comando de voz seria “table copy”, para copiar na projeção o comando seria “Kinect copy”. Com este sistema as aplicações sabem quando o comando é direcionado a elas. Se tiver o nome correto o comando é aceite senão é ignorado.

Para ilustrar o funcionamento desta arquitetura e para se perceber melhor o fluxo de trabalho aqui apresentado será dado um exemplo do que acontece quando o utilizador quer copiar um objeto na mesa interativa e de seguida quer enviar a cópia para a outra superfície.

Para copiar um objeto utilizando a voz o utilizador coloca o dedo em cima do objeto e diz: “table copy”. Nesse momento o Kinect capta o comando de voz para a aplicação C#. A aplicação reconhece o comando e envia-o para o servidor que verifica que recebe um comando de voz. Para que não exista uma restrição no número de dispositivos ligados ao sistema, o servidor envia um broadcast do comando de voz recebido, sem que haja filtragem e é trabalho do cliente o de descartar ou não a mensagem, verificando assim se esta lhe foi dirigida.

A aplicação “Web App sem toque” verifica se o comando tem o seu nome. Como não tem então ignora-o. Ao mesmo tempo a aplicação da mesa – “Web App com toque” - verifica se o comando tem o seu nome, vê que sim e aceita o comando. Logo depois é identificado o objeto que o utilizador está a agarrar/tocar e é feita a cópia desse objeto.

No passo a seguir o utilizador envia a cópia para a projeção, através de um swipe para cima. A mesa reconhece que o gesto é de envio e envia a informação ao servidor nodejs, que trata também da troca de informação entre superfícies (componente representada entre as aplicações na imagem 3.8). O servidor recebe a informação e passa o menu para a aplicação sem toque. Esta exhibe o menu recebido, na projeção.

3.6 Funcionalidades

Como já foi referido na secção 3.2.2 deste capítulo as ações que são suportadas para a aplicação são: copiar, minimizar, maximizar, sincronizar, fechar/apagar e enviar para a superfície contrária (troca de objetos entre superfícies). Os objetos podem ser movidos para qualquer ponto da superfície que se está a trabalhar. Estas funcionalidades foram contextualizadas numa aplicação para escolher um destino de férias, pois é algo que capta a atenção dos participantes nos testes e também é um tema em que faz sentido ter este tipo de ações. Todas as funcionalidades têm três formas de serem executadas: utilizando os botões contidos nos objetos, utilizando apenas gestos, e utilizando comandos de voz em adição aos gestos.



Figura 3.9: Cópia de um objeto.

A função copiar, como o próprio nome indica, faz uma cópia do objeto, ficando este no mesmo estado que o objeto original (Figura 3.9). Esta funcionalidade funciona de igual modo quer na projeção quer na mesa interativa.

A função de minimizar esconde o objeto criando mais espaço livre para o utilizador trabalhar sem que este necessite de o eliminar/fechar. O objeto pode ser escondido/minimizado para baixo (Figura 3.10b) e para o lado esquerdo (Figura 3.10a) ou direito (Figura 3.10c). Em ambas as superfícies a ação tem comportamento igual. Quando um objeto é minimizado este fica apenas com um terço à mostra para o utilizador perceber que ainda ali está um objeto. Da mesma maneira, depois de minimizar, é possível restaurar o objeto para voltar a ver a informação contida no objeto minimizado/escondido.



(a) Minimizar esquerda.



(b) Minimizar baixo.



(c) Minimizar direita

Figura 3.10: Possíveis formas de minimizar

A sincronização permite que dois ou mais objetos do mesmo tipo mostrem a mesma informação (Figura 3.11). Tudo o que é alterado num objeto é também alterado no outro, ao mesmo tempo.

Em todos os objetos é permitida a opção de fechar e a opção de enviar para a superfície contrária. Se o utilizador estiver a trabalhar na mesa pode querer enviar um objeto para a projeção e controlá-lo através de gestos sem ajuda do toque. O objetivo deste protótipo é apenas auxiliar num estudo de interação gestual. Assim a informação mostrada não é dinâmica nem consultada na Internet. É uma aplicação com informação estática limitada à zona metropolitana de Lisboa.



Figura 3.11: Sincronização de dois objetos.

Capítulo 4

Estudos de Interação

No capítulo anterior foi apresentado um processo de desenho de aplicações multimodais. Neste capítulo é apresentada a concretização de duas etapas desse processo. A primeira, que se adapta à etapa de compreensão da interação em ambientes sem restrições tecnológicos, corresponde ao primeiro estudo efetuado; e a segunda, que se adapta à validação da interação com protótipo de alta-fidelidade corresponde ao segundo estudo.

A primeira secção mostra o primeiro estudo efetuado, cujo principal objetivo foi recolher os comandos de gestos e voz, para determinadas ações, em aplicações multimodais e multisuperfície. É apresentado um enquadramento do estudo, as tarefas dadas aos participantes, qual o procedimento seguido, os resultados e por fim a discussão dos resultados obtidos. Este estudo permitiu realizar um segundo estudo que está descrito na secção seguinte.

A segunda secção mostra o segundo estudo feito neste trabalho. O objetivo deste estudo é perceber qual o efeito de alguns fatores na interação em aplicações multimodais, multiutilizador e multisuperfície. Estes fatores foram: o número de pessoas que estão a interagir com a aplicação e o comportamento desta, ou seja, alterar o aparecimento de novos objetos e alternar a superfície onde os objetos são mostrados.

É de salientar que foi introduzido neste segundo estudo suporte para múltiplos utilizadores de forma a avaliar o impacto que existe na interação multimodal e multisuperfície.

4.1 Primeiro Estudo

4.1.1 Enquadramento e preparação do estudo

O objetivo principal deste estudo foi perceber quais os gestos e os comandos de voz que eram utilizados em várias ações, em cada uma das superfícies. O resultado deste estudo permitiu recolher uma lista de gestos e comandos de voz para a construção de um segundo protótipo, onde este resultado foi aplicado.

Para este fim construiu-se um protótipo muito simples em que foram apresentados apenas objetos e foi pedido aos participantes para realizarem determinadas tarefas. Durante

a realização das tarefas os participantes tiveram total liberdade de escolhas dos gestos e comandos de voz.

Para cada objeto foi permitido realizar ações através de botões. Estes botões permitiam efectuar as mesmas acções que se conseguiam executar através de gestos e voz. Foi pedido aos participantes para verificar se o desenho do botão ilustrava, de forma clara, a ação correspondente.

4.1.2 Tarefas

A primeira tarefa dada aos participantes foi perceber o que os botões contidos no objeto ilustravam, ou seja, as pessoas olhavam para cada botão e comentavam a sua função. A figura 4.1 mostra os botões que foram apresentados aos participantes.



Figura 4.1: Botões apresentados aos participantes.

Após esta avaliação foram efetuados dois estudos diferentes. Em primeiro lugar foi pedido aos participantes para realizarem as ações sobre o objeto utilizando apenas gestos, em cada uma das superfícies. Em seguida foi pedido para efetuar as mesmas ações mas utilizando gestos com a adição de comandos de voz.

As ações a executar para todos os objetos da aplicação foram as seguintes: (1) transferir objetos de uma superfície para a outra; (2) sincronizar objetos; (3) esconder/minimizar um objeto; (4) copiar objetos e por fim (5) mover os objetos de sítio. Este estudo foi efetuado na mesa interativa e de seguida utilizando a projeção. Os participantes mostravam como efetuavam as ações na mesa interativa e de seguida era feito o mesmo procedimento utilizando a projeção, com a diferença que os gestos eram efetuados sem a ajuda do toque.

4.1.3 Participantes

Nesta sessão experimental participaram 19 pessoas, 4 do sexo feminino e 15 do sexo masculino, com uma média de 25 anos. Todos os participantes eram estudantes universitários e 18 deles afirmaram ter experiência em interação gestual com toque em dispositivos como tablets e smartphones. Apenas 1 utilizador afirmou ter experiência com mesas interativas. Dos 19 participantes, 14 assumiram ter experiência em interação gestual sem toque, em dispositivos como Nintendo Wii, PS3 Move e realidade virtual, sendo que destas,

3 pessoas já tinham utilizado o Kinect. Em relação aos comandos de voz, 12 participantes assumiram ter experiência com este tipo de interação em aplicações de telemóveis.

4.1.4 Procedimento

O estudo iniciou-se com uma explicação do funcionamento da aplicação e qual o seu propósito. De seguida recolheu-se o perfil de cada participante recorrendo a um questionário.

Durante a experiência, os participantes tiveram total liberdade de interação gestual e de voz, ou seja, não foram impostas quaisquer condições ou restrições ao uso de gestos e comandos de voz, pelos participantes.

Cada pessoa participou numa sessão do estudo individualmente, manipulando os objetos em duas superfícies distintas, a mesa interativa e a projeção. Em ambas as superfícies foi permitido o uso de comandos de voz e apenas os gestos eram diferentes no que toca à existência de toque. No início da sessão foi mostrado a cada participante os botões embutidos em cada objeto da aplicação e, sem dizer as ações que eram permitidas, foi pedido às pessoas para ligarem um botão à ação que lhes parecesse mais adequada. Se o participante não conseguisse perceber ou se não conseguisse fazer corresponder a ação correta, então ser-lhe-ia perguntado qual o desenho gráfico, que ilustraria melhor o que era pretendido. De seguida foi pedido aos participantes que sugerissem um gesto e um comando de voz para cada ação pedida. Por exemplo, foi pedido para efectuar uma cópia do objeto, primeiro utilizando apenas os gestos e de seguida utilizando, em adição aos gestos, os comandos de voz. Neste último caso o participante poderia escolher utilizar só gestos ou utilizar só comandos de voz ou utilizar ambos.

4.1.5 Resultados

Como cada sessão de estudo era constituída por três fases, a apresentação dos resultados encontra-se dividida em três partes: na primeira parte, a opinião dos participantes referentes aos botões dos objetos da aplicação; na segunda parte, quais os gestos efetuados num cenário de interação puramente gestual; por último, na terceira parte, quais os gestos e comandos de voz utilizados numa interação gestual com o auxílio de comandos de voz.

Representação gráfica das ações

Na tabela 4.1 são apresentados os botões mostrados aos participantes e é mostrado o número de pessoas que perceberam o seu significado. Na coluna NOTAS é mostrado o que foi percebido pelos participantes que não entenderam para que era usado aquele botão e algumas sugestões dadas pelos mesmos.






Botão	Função	Percebeu sua Função	Não percebeu Função	Notas
	Minimizar	18	1	O participante pensou que servia para diminuir o menu
	Maximizar	19	0	—
	Copiar	15	4	Sugestão: Colocar um sinal '+' entre os dois menus
	Sincronizar	5	14	Oito participantes pensavam que este botão servia para refresh/reload Dois participantes pensaram que o botão era Switch Dois participantes pensaram que era para rodar Dois participantes pensaram que era para voltar a fechar o menu
	Trocar de Superfície	11	8	Sugestão: Colocar um ecrã em cima da seta

Tabela 4.1: Resultados obtidos sobre a representação gráfica dos botões.

Interação apenas gestual

Nesta sessão do estudo foi pedido aos participantes para utilizarem apenas gestos. Na tabela 4.2 são mostrados os gestos efetuados na mesa interativa para cada ação pedida e na tabela 4.3 pode-se observar os gestos efetuados na projeção, utilizando gestos sem a ajuda do toque.

Interação gestual com o auxílio de comandos de voz

Na última sessão foi pedido aos participantes para efetuarem as mesmas tarefas mas agora podendo utilizar os comandos de voz. Na tabela 4.4 é mostrado o resultado para a mesa interativa e na tabela 4.5 é mostrado o resultado da projeção.

4.1.6 Análise e Discussão dos resultados

Na primeira parte dos testes, onde foi pedido aos participantes para tentarem perceber quais as funções dos botões, ficou claro que os botões de minimizar e maximizar ilustram bem a sua função, principalmente devido ao facto dos botões terem sido baseados no aspecto gráfico do sistema operativo Microsoft Windows, que apresenta um alto nível de utilização. Os outros três botões, copiar, sincronizar e trocar de superfícies, precisaram de sofrer alterações para melhorar a sua representação gráfica, para que não houvesse dúvidas de qual a ação que o botão executava. Neste aperfeiçoamento foram tidas em conta as sugestões dadas pelos participantes.

Depois destes testes foram feitas melhorias na apresentação gráfica não só dos botões mas também dos objetos. A figura 4.2 ilustra bem a mudança que os objetos e botões

Função	Gesto efetuado	Nº de Participantes
Minimizar	Swipe para baixo	19
Maximizar	Puxar o objeto para cima	19
Copiar	Dois toques em cima do objeto	15
	Um dedo em cima de um menu e dois dedos na superfície	1
	Área de dropdown onde se mete o original e depois fica lá a cópia	1
	Dois dedos de uma mão seguram um menu e outro puxa uma cópia	1
	Não sabe	1
Sincronizar	Sobrepor objetos	13
	Gesto circular em torno de dois objetos	1
	Duplo clique (ao mesmo tempo) em dois objetos do mesmo tipo	5
Trocar objeto entre superfícies	Swipe para cima	19
Limpar área de trabalho	Swipe na diagonal com uma mão	11
	Usando as duas mãos e afastando uma da outra	3
	Gesto de limpar a mesa	5

Tabela 4.2: Tabela de gestos efetuados na mesa interativa.

sofreram.

Na segunda parte dos testes, onde os participantes faziam as ações utilizando apenas os gestos, podemos observar que na mesa interativa as ações de minimizar, maximizar e trocar entre superfícies, tiveram um grande consenso em relação ao gesto utilizado, não havendo qualquer dúvida que os gestos feitos para minimizar, maximizar e trocar de superfície são os mais indicados para estas ações. As restantes ações causaram mais

Função	Gesto efetuado	Nº de Participantes
Minimizar	Swipe para cima	19
Maximizar	Puxar o objeto para baixo	19
Copiar	Dois toques em cima do objecto (simula dois toques no ar)	6
	Arrasta o objeto para uma área de dropdown (a cópia fica dentro dessa área)	
	Agarrar um menu com uma mão e com a outra puxar menu (similar ao zoom in)	9
	Não sabe	3
Sincronizar	Sobrepor objetos	8
	Duplo toque (ao mesmo tempo) em dois objetos do mesmo tipo (simula no ar)	4
	Não sabe	7
Trocar objeto entre superfícies	Swipe para baixo	19
Limpar área de trabalho	Não sabe	19
Agarrar um objeto	Fechar a mão	19
Largar um objeto	Abrir a mão	19

Tabela 4.3: Tabela de gestos efetuados na projeção.

dúvidas nos participantes havendo várias formas de fazer a mesma ação.

Na superfície de projeção era necessário perceber como o participante iria agarrar um objeto, e como se pode observar na tabela 4.3, não existe dúvida que para agarrar se fecha a mão e para largar basta abrir a mão. Esta é uma situação que impede a aplicação do

Função	Gesto e/ou comando de voz	Nº de Participantes
Minimizar	Swipe para baixo	2
	Tocar no objeto e dizer 'hide'	16
	Tocar no objeto e dizer 'minimize'	1
Maximizar	Tocar no objeto e dizer 'show'	14
	Puxar o objeto	3
	Tocar no objeto e dizer 'reveal'	1
	Tocar no objeto e dizer 'restore'	1
Copiar	Sem tocar na superfície, dizer o nome do objeto seguido de 'copy'	5
	Tocar no objeto e dizer 'copy'	14
Sincronizar	Tocar em dois objetos e dizer 'Sync' ou 'Sincronize'	19
Trocar objeto entre superfícies	Swipe para cima	3
	Tocar o objeto e dizer 'send'	16
Limpar área de trabalho	Gesto de Limpar	2
	Dizer apenas 'clean' ou 'restart'	17

Tabela 4.4: Tabela de gestos e comandos de voz dados na mesa interativa.

processo de desenho como se desejaria. É uma restrição técnica uma vez que o SDK do Kinect não consegue reconhecer a mão aberta ou fechada. Esta restrição impede que os gestos de agarrar e largar um objeto sejam efetuados naturalmente. Para ultrapassar este problema recorreu-se aos comandos de voz: os participantes para agarrar um objeto davam o comando de 'DRAG' e para largar davam o comando de 'DROP'.

Como aconteceu na mesa interativa, na superfície de projeção também se identificou

Função	Gesto efetuado	Nº de Participantes
Minimizar	Swipe para cima	2
	Agarrar no objeto e dizer ‘hide’	16
	Agarrar no objeto e dizer ‘minimize’	1
Maximizar	Agarrar no objeto e dizer ‘show’	14
	Puxar o objeto	3
	Agarrar no objeto e dizer ‘reveal’	1
	Agarrar no objeto e dizer ‘restore’	1
Copiar	Sem tocar na superfície, dizer o nome do objeto seguido de ‘copy’	5
	Agarrar no objeto e dizer ‘copy’	14
Sincronizar	Agarrar em dois objetos e dizer ‘Sync’ ou ‘Sincronize’	19
Trocar objeto entre superfícies	Swipe para baixo	3
	Agarrar o objeto e dizer ‘send’	16
Limpar área de trabalho	Dizer apenas ‘clean’ ou ‘restart’	19
Agarrar um objeto	Fechar a mão	19
Largar um objeto	Abrir a mão	19

Tabela 4.5: Tabela de gestos efetuados e comandos de voz dados na projeção.



Figura 4.2: Menu da aplicação depois da melhoria.

que não houve muita dúvida em relação às ações de minimizar, maximizar e troca de superfícies. Nas restantes ações os participantes mostraram estar mais confusos e houve até algumas pessoas que não souberam fazer o gesto para determinada ação.

Na última parte dos testes, onde se utilizaram os comandos de voz, é de notar que não houve tanta diferença do número de comandos dados para a mesma ação, como quando se utilizaram apenas os gestos. Quando a utilização da voz não era permitida, os participantes não eram tão consensuais na sua escolha, ou seja, existiam vários gestos para uma mesma ação. Com a adição dos comandos de voz este problema deixou de existir uma vez que a maioria dos participantes fez a mesma ação (em termos de gesto), alterando apenas o comando de voz que era dado. Um aspeto interessante que foi observado, foi o tempo de resposta ao que foi pedido, uma vez que quando foram utilizados apenas gestos, os utilizadores demoraram mais tempo a dar uma resposta e alguns afirmaram não fazer ideia de como fazer o que foi pedido, ao passo que quando se possibilitou a utilização de comandos de voz, notou-se uma diminuição do tempo de resposta. Este resultado pode ter sido influenciado por duas causas: a primeira foi a dificuldade de chegar a uma representação somente gestual para as ações que não envolvem manipulação direta, que é resolvida pela possibilidade de dar comandos verbalmente e assim conseguiam chegar a um resultado mais rapidamente, e a segunda pode ter sido devido ao facto de os participantes já terem pensado naquela ação anteriormente quando executaram o teste utilizando apenas gestos.

Com os resultados obtidos podemos concluir que para aplicações multimodais e multisuperfície, é altamente recomendado o uso de gestos com a junção de comandos de voz, uma vez que torna mais fácil a desambiguação de gestos e de manipular objetos mais facilmente quando se quer efetuar uma ação mais abstrata sobre eles.

Os gestos e os comandos de voz que foram obtidos com este estudo foram implementados no protótipo final da aplicação. Os gestos que foram repetidos mais vezes pelos participantes e os todos os comandos de voz identificados neste estudo foram utilizados para melhorar a interação com o protótipo final.

4.2 Segundo Estudo

4.2.1 Enquadramento e preparação do estudo

O objetivo principal deste estudo foi perceber de que forma as três variáveis em estudo, - número de pessoas em torno da aplicação, superfícies em que os objetos copiados aparecem e superfícies em que os objetos novos são mostrados - influencia no fluxo de trabalho e no modo de interação com este tipo de aplicações. Para este fim desenhou-se uma experiência em que os participantes podem recorrer a uma junção de interações: interação por gestos com e sem toque com o auxílio dos comandos de voz utilizando mais do que uma superfície como ambiente de trabalho.

Para este fim foi construído um protótipo de uma aplicação web com o intuito de permitir aos participantes planejar uma ida de férias a uma cidade à escolha. Para isso podiam ver os locais turísticos dessa cidade, fotografias, os hotéis e os restaurantes. Também era permitida a marcação de voos para essa cidade.

Foram dadas várias tarefas aos participantes e durante estas, foi-lhes permitido utilizar apenas os gestos e comandos de voz que foram compilados na sessão experimental anterior, sem que estes lhes tivessem sido previamente indicados.

As ações que poderiam ser executadas sobre os objetos poderiam ser feitas de 3 formas: (1) utilizando os botões, quer através de toque na superfície interativa, quer apontando para eles na superfície de projeção; (2) utilizando apenas gestos e por fim (3) utilizando gestos e comandos de voz.

Para obter os dados para análise foi feito um script que permitiu a recolha dos dados necessários referentes à interação. A informação conseguida mostrava quais as ações executadas, quantas vezes uma ação era feita e a forma como essa ação era conseguida, se utilizando botões, ou gestos ou com a ajuda de comandos de voz.

4.2.2 Tarefas

Foi elaborado um conjunto de tarefas com o propósito de obrigar os participantes do teste a usar todas as funcionalidades da aplicação. Cada tarefa foi feita de modo a que houvesse um seguimento da tarefa anterior, assim a utilização da aplicação fazia sentido para o participante ou participantes no teste.

As tarefas são ilustradas com figuras que mostram os participantes na execução de algumas tarefas. As tarefas efetuadas por todos os participantes foram:

- Escolher a cidade de Lisboa como destino de férias, ver fotografias da cidade e escolher duas fotografias (por pessoa) que chamaram mais à atenção. (figura 4.3)
- Depois de ver as fotos de Lisboa, saber quais os locais turísticos que a cidade oferece e escolher três como prioridade de visita.
- Encontrar o hotel de três estrelas, mais barato, na zona do Saldanha.



Figura 4.3: Participante a executar a primeira tarefa.

- Encontrar o hotel mais barato de três zonas: Sete Rios, Cidade Universitária e Sal-danha.
- Depois de encontrado o hotel mais barato ver quais os três restaurantes mais perto desse hotel.
- Escolher o restaurante com os amigos ou família que por motivos de espaço não conseguem estar perto da mesa interativa.¹ (figura 4.4)



Figura 4.4: Participante a interagir com a projeção.

- Depois de escolhido o hotel mais barato e o restaurante que mais agradou a todos, pediu-se para marcarem um voo para cada pessoa para o dia 10 de Dezembro de 2012.

Para cada tarefa o estado do ambiente de trabalho era o que ficava com o final da tarefa anterior, ou seja, os participantes no início de cada tarefa escolhiam o que fazer

¹Para esta tarefa encontravam-se, uma ou mais, pessoas afastadas da mesa interativa que desempenhavam o papel de familiar ou amigo.

com o ambiente em que se encontravam, excetuando a primeira tarefa que começava com o menu inicial imposto.

Para as tarefas que necessitavam de encontrar locais específicos, os participantes tinham à sua disposição um objecto mapa (ver figura 4.5), que permitia a interação gestual.



Figura 4.5: Exemplo de um cenário da aplicação.

4.2.3 Participantes

Nesta sessão participaram 36 pessoas, 11 do sexo feminino e 25 do sexo masculino, com idades compreendidas entre os 18 e os 30 anos. 34 dos participantes assumiram ter experiência com dispositivos de interação por toque e apenas 10 com experiência em interação gestual sem toque. Em relação a interação por voz apenas 8 participantes tinham experiência com este tipo de interação. É de salientar que, nenhum dos participantes que esteve presente no primeiro estudo, participou neste segundo estudo.

4.2.4 Procedimento

Para a realização dos testes a superfície de projeção foi colocada à frente da mesa interativa (ver figura 4.6), assim os participantes tinham uma boa visão das duas superfícies de trabalho.

Para efetuar este estudo foram construídas 6 rondas de testes. Em cada ronda variaram-se as variáveis em estudo, reflectindo-se isso também na forma como a aplicação se comportava. Na aplicação existiam os conceitos de menus (figura 4.7 (1)) e cópias (figura 4.7

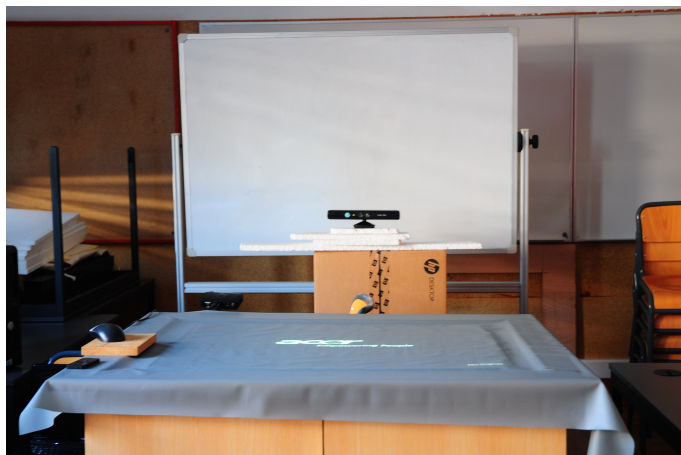


Figura 4.6: Ambiente físico da aplicação

(2)). Os menus são aqueles que permitem escolher o que se quer ver a seguir, as cópias são os elementos que foram copiados, podem ser menus ou imagens. É com estes dois tipos de objetos que o comportamento da aplicação foi alterado.



Figura 4.7: (1) Exemplo de menus que a aplicação tem. (2) As cópias feitas pela aplicação

Na primeira ronda todos os objetos eram mostrados na mesa interativa e quando era feita uma cópia esta aparecia na mesma superfície que o original. Na segunda ronda era alterado o comportamento das cópias, que abriam na superfície contrária ao seu original. Na terceira e quarta ronda os objetos eram todos apresentados na projeção e o comportamento das cópias era diferente: na terceira ronda as cópias abriam na mesma superfície que o original e na quarta ronda abriam na superfície contrária. Na quinta e sexta ronda de testes o primeiro menu era apresentado na mesa interativa mas os seguintes dependiam da superfície onde se encontrava o seu menu antecessor. Por exemplo, se o participante enviasse o primeiro menu para a projeção, o menu seguinte, que derivava deste primeiro,

era aberto na projeção. Se o primeiro menu fosse mantido na mesa interativa então o segundo menu abria na mesa, ou seja, os objetos eram abertos ou chamados na mesma superfície em que o seu “pai” se encontrava. Em relação às cópias, o comportamento era igual ao referido anteriormente: na quinta ronda as cópias abriam na mesma superfície que o original e na sexta ronda abriam na superfície contrária. Em qualquer ronda foram efetuados testes com grupos de um, dois e três participantes. A tabela 4.6 resume como se variou o local de aparecimento de cópias e novos objetos nas 6 rondas.

Rondas	Onde aparecem cópias	Apresentação dos objectos
1 ^a	Mesma superfície que o original	Mesa interativa
2 ^a	Superfície contrária ao original	Mesa interativa
3 ^a	Mesma superfície que o original	Projeção
4 ^a	Superfície contrária ao original	Projeção
5 ^a	Mesma superfície que o original	Onde está o objeto 'PAI'
6 ^a	Superfície contrária ao original	Onde está o objeto 'PAI'

Tabela 4.6: Explicação dos factores variados nas 6 rondas de teste.

O estudo iniciou-se com a explicação do tipo de interação e qual o objetivo da aplicação: escolher uma cidade para ir de férias.

Cada ronda contou com três grupos de testes: um de uma pessoa, outro com duas pessoas e outro com três pessoas. Foram dadas as mesmas tarefas a todos os grupos de todas as rondas e estas foram pensadas de modo a englobar todas as ações e de modo a que os participantes fossem obrigados a utilizar as duas superfícies de trabalho.

No final de cada sessão, cada participante preencheu um questionário com toda a informação sobre a experiência com este tipo de interação e com questões sobre o comportamento da aplicação, indicando se este comportamento lhe parecia normal ou não e se este tipo de interação trazia alguma vantagem ao ser utilizado. Cada questionário era adaptado a cada ronda de testes, em função do comportamento da aplicação nessa ronda. No anexo A encontra-se o questionário apresentado aos participantes da primeira ronda. As perguntas Q5 e Q6 eram adaptadas a cada ronda de testes.

4.2.5 Análise dos Resultados

Nesta secção serão apresentados e analisados todos os dados obtidos depois dos dezoito testes efetuados. Foram analisadas todas as variáveis referidas anteriormente.

A primeira variável a ter em conta é o número de participantes em torno da aplicação, que poderão ser um, dois ou três. Na imagem 4.8 podemos ver o gráfico que mostra o efeito do número de pessoas nos comandos que são dirigidos a cada superfície de interação. Assim, é possível perceber se este número altera ou não o número de comandos de qualquer modalidade em cada superfície.

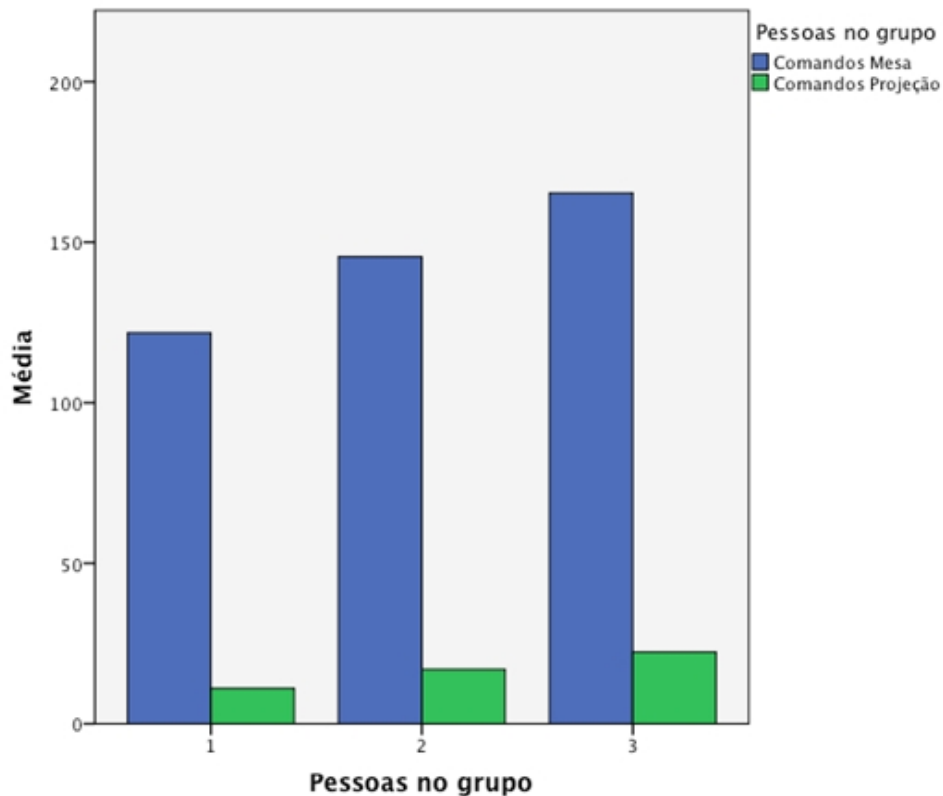


Figura 4.8: Efeito do número de pessoas na superfície de interação.

Para fazer esta análise foi utilizada a técnica de análise multivariada de variância – MANOVA, em que o resultado foi, $F(5,28) = 22,495$, $P < 0,0005$, Wilk's $\lambda = 0,03$, $\epsilon^2 = 0,828$.

Com este resultado, foi possível observar que existe uma diferença estatística significativa no número de comandos que são enviados para a mesa ou para a superfície de projeção em função do número de pessoas que estão a interagir com a aplicação. Os resultados obtidos dos testes dos efeitos entre sujeitos, permitem concluir que o número de pessoas no grupo têm efeito estatisticamente significativo quer no número de comandos dado na mesa, $F(3,15) = 150,74$, $P < 0,0005$, quer no número de comandos na superfície de projeção, $F(3,15) = 16,84$, $P < 0,0005$.

De seguida, foram efetuados testes de Tukey's HSD que permitiram identificar se existiam diferenças para os comandos na mesa e na projeção entre os vários grupos de pessoas. Na mesa, o resultado dos testes entre os grupos de uma e três pessoas é de $P=0,05$, de uma e duas pessoas é de $P=0,359$ e entre os grupos de duas e três pessoas é de $P=0,48$. O mesmo foi efetuado para a superfície de projeção em que a diferença entre os grupos de uma e duas pessoas é de $P=0,588$, de uma e três pessoas é de $P=0,176$ e de duas e três pessoas é de $P=0,655$.

Analisando os resultados obtidos verificamos que existe uma diferença significativa de $P=0,05$ para os comandos realizados na mesa entre os grupos de uma e três pessoas, ou

seja, o número de comandos efectuados na mesa no grupo de uma pessoa é menor do que no grupo de três pessoas, apesar de se pedir para executar exatamente as mesmas tarefas.

A segunda observação (gráfico 4.9) é sobre o efeito do número de pessoas nas modalidades usadas. Para este resultado, como no teste anterior, foi utilizado o teste MANOVA em que o resultado foi o seguinte: $F(8,45) = 15,123$, $P < 0,0005$, Wilk's $\lambda = 0,017$, $\epsilon^2 = 0,741$.

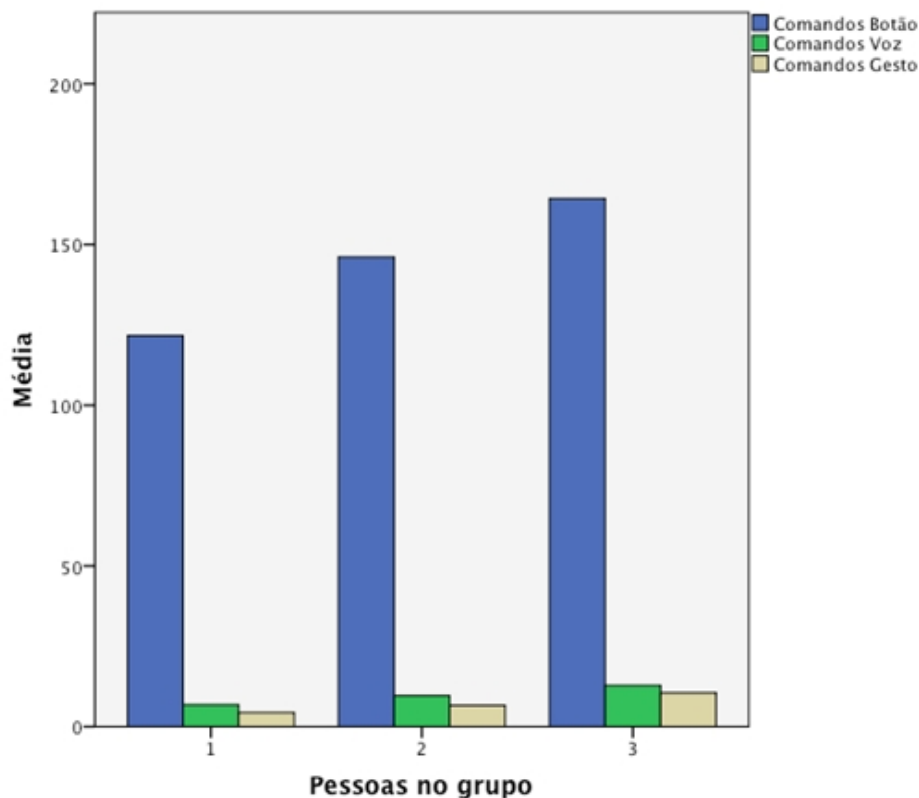


Figura 4.9: Efeito do número de pessoas nas modalidades de interação usadas.

Os resultados obtidos dos testes dos efeitos entre sujeitos permitem concluir que o número de pessoas no grupo têm efeito estatisticamente significativo quer no número de comandos dados por seleção de botão: $F(3,15) = 125,563$, $P < 0,0005$, no número de comandos de voz: $F(3,15) = 10,822$, $P < 0,0005$, ou no número de comandos por gesto: $F(3,15) = 40,605$, $P < 0,0005$.

Este resultado permite concluir que o número de pessoas em torno da aplicação tem impacto sobre as modalidades utilizadas para interagir com a aplicação, ou seja, quantas mais pessoas houver em torno da mesa maior é o número de comandos em todas as modalidades. De referir, novamente, que as tarefas eram exatamente as mesmas independentemente do número de pessoas no grupo.

Seguindo estes testes significativos com testes Tukey's HSD foram identificadas as diferenças entre o número de comandos dados por seleção de botões, o número de comandos por voz e o número de comandos por gesto, tendo em conta os diferentes grupos

de pessoas.

A diferença para os comandos dados por seleção de botão entre os grupos de uma e duas pessoas foi de $P = 0,397$, entre os grupos de uma e três pessoas foi de $P = 0,082$ e entre os grupos de duas e três pessoas foi de $P = 0,593$. A diferença para os comandos de voz entre o grupo de uma e duas pessoas foi de $P = 0,793$, entre os grupos de uma e três pessoas foi de $P = 0,373$ e entre duas e três pessoas foi de $P = 0,749$. Por último, as diferenças para os comandos por gestos entre os grupos de uma e três pessoas foi de $P = 0,006$, entre os grupos de uma e duas pessoas foi de $P = 0,374$ e entre os grupos de duas e três pessoas foi de $P = 0,091$.

Assim concluímos que só existe uma diferença significativa entre os grupos de uma e três pessoas no que toca ao número de comandos dados por gesto. Então, quando o número de pessoas aumenta, o número de comandos por gesto também aumenta, causando um impacto na interação entre os grupos de uma e três pessoas.

A próxima variável identificada foi o impacto da função da abertura de cópias. O gráfico que ilustra este teste é apresentado na figura 4.10. Através do teste χ^2 foi identificada a diferença na satisfação com a funcionalidade das cópias em função destas aparecerem na mesma superfície ou na superfície contrária do objeto original. O resultado obtido foi $\chi^2(1,36) = 18$, $P < 0,0005$.

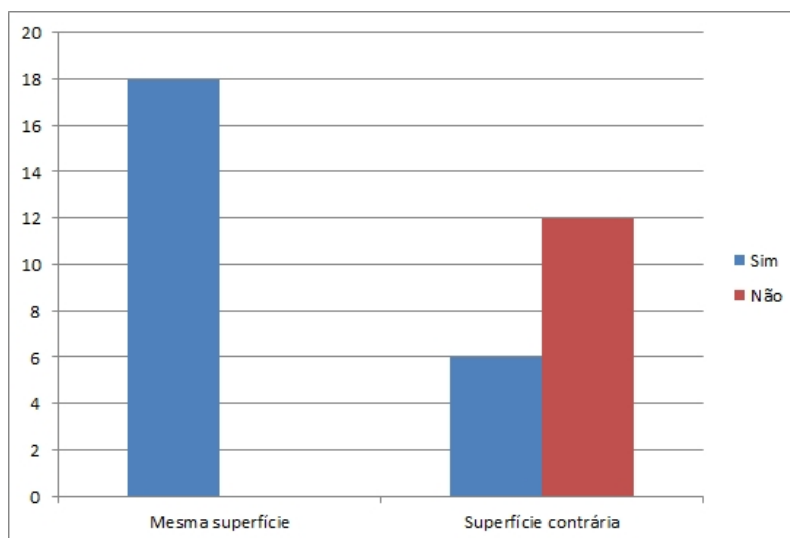


Figura 4.10: Efeito da abertura da cópia na satisfação das pessoas.

Com o estudo do efeito da abertura da cópia na satisfação com a funcionalidade é possível verificar, através da análise do gráfico 4.10, que quando as cópias abrem na mesma superfície que o original, os participantes mostram-se plenamente satisfeitos mas o mesmo não acontece quando as cópias abrem na superfície contrária ao original, pois existe um número de não satisfação muito grande em relação ao de satisfação. O teste χ^2 mostra que existe uma diferença significativa ($P < 0,0005$) na satisfação com a funcionalidade das cópias em função de elas aparecerem na mesma superfície ou na superfície

contrária ao original.

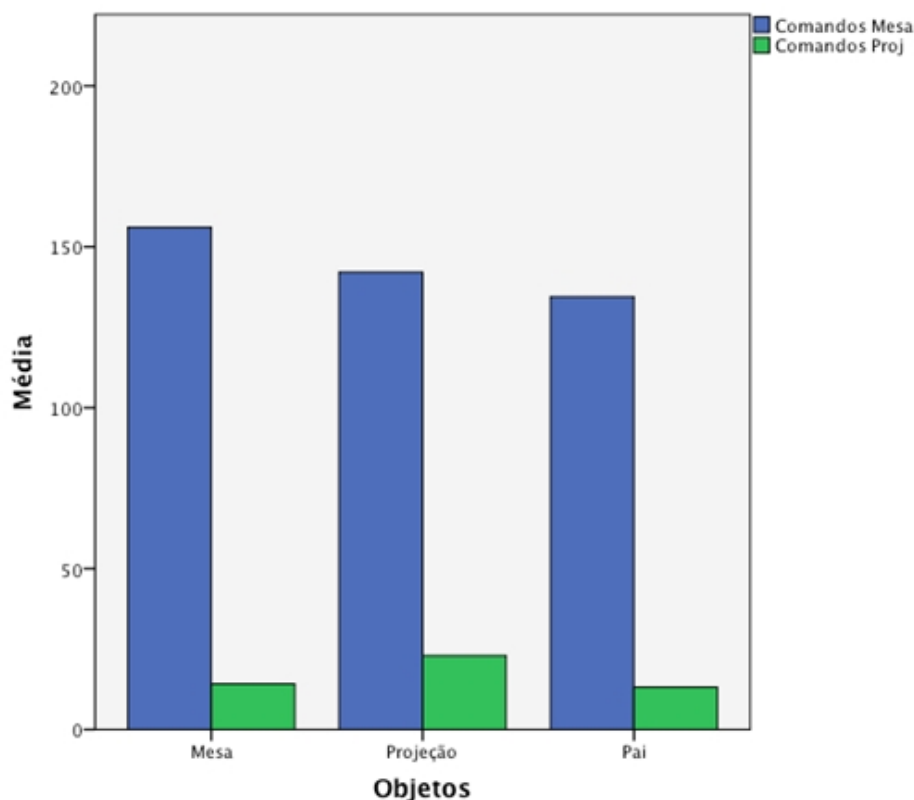


Figura 4.11: Efeito do local onde aparecem os objetos no uso das superfícies de interação.

Em seguida procedemos à análise da ultima variável em estudo, a diferença de superfícies onde os objectos aparecem. Foi analisado se o local onde o objeto aparece condicionava o uso das superfícies de interação (ver gráfico 4.11) e se condicionava o uso das modalidades utilizadas (ver gráfico 4.12). Para o efeito do local onde aparecem os objetos no uso das superfícies de interação temos um resultado, com os testes de MANOVA, de: $F(5,28) = 1,344$, $P=0,278$, Wilk's $\lambda = 0,704$, $\epsilon^2 = 0,161$. O efeito do local onde aparecem os objectos nas modalidades utilizadas temos um resultados dos testes de MANOVA de: $F(5,26) = 1,073$, $P=0,404$, Wilk's $\lambda = 0,643$, $\epsilon^2 = 0,198$.

Com a análise destes resultados pode observar-se que não existem diferenças estatisticamente significativas, para se poder verificar alguma influência sobre este tipo de aplicações.

Analisando o gráfico 4.13 consegue-se perceber que os participantes, de forma geral, não se importam com a superfície em que os objectos aparecem, apesar de se notar que quando estes estão a aparecer na projeção há mais variação na opinião dos participantes. Onde se observou uma satisfação total foi quando os objectos abriam onde estava o seu antecessor, ou seja, aparecia na superfície onde estava o objecto pai.

Através do teste χ^2 não se verificou uma diferença significativa na satisfação com o local onde surgem objetos novos em função de estes aparecerem sempre na mesa inte-

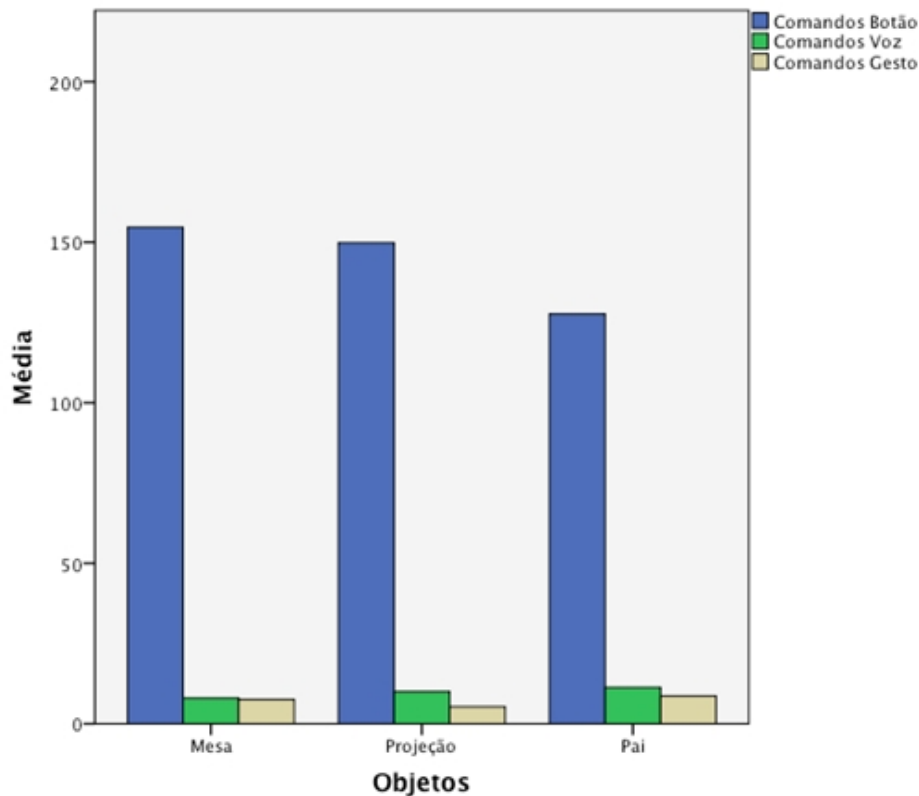


Figura 4.12: Efeito do local onde aparecem os objetos nas modalidades usadas.

rativa, na superfície de projeção, ou na mesma superfície onde está o objeto que lhes dá origem, $\chi^2(2,36)=3,938$, $P=0,14$.

De seguida é mostrado o gráfico de satisfação dos participantes com as componentes de interação (figura 4.14), conseguido através das respostas aos questionários feitos no fim de cada teste. No questionário efetuado era perguntado aos participantes se os comandos de voz e os gestos suportados pela aplicação correspondiam ao esperado pelos participantes. Foi também avaliada a satisfação dos utilizadores em terem usado este tipo de interação e se no futuro gostariam de a voltar a utilizar. A escala de satisfação dada era de 1 a 5, sendo o valor mais baixo um grau de plena insatisfação e o grau mais alto a satisfação máxima.

Como se pode ver pela figura 4.14, tanto os comandos por gesto, como os comandos por voz, foram avaliados positivamente pelos participantes nos testes. No entanto, os comandos por voz tiveram uma avaliação mais positiva, o que pode indiciar uma preferência por este tipo de interação. No geral, os utilizadores ficaram satisfeitos com a utilização da aplicação, querendo no futuro voltar a usar este tipo de interação.

O teste de Kruskal-Wallis, para verificar o efeito do número das pessoas no grupo das métricas de satisfação, não detetou qualquer efeito estatisticamente significativo do número de pessoas no grupo quer para a satisfação com os comandos de voz, a satisfação com os gestos, a satisfação geral, ou a vontade de usar um sistema semelhante no futuro.

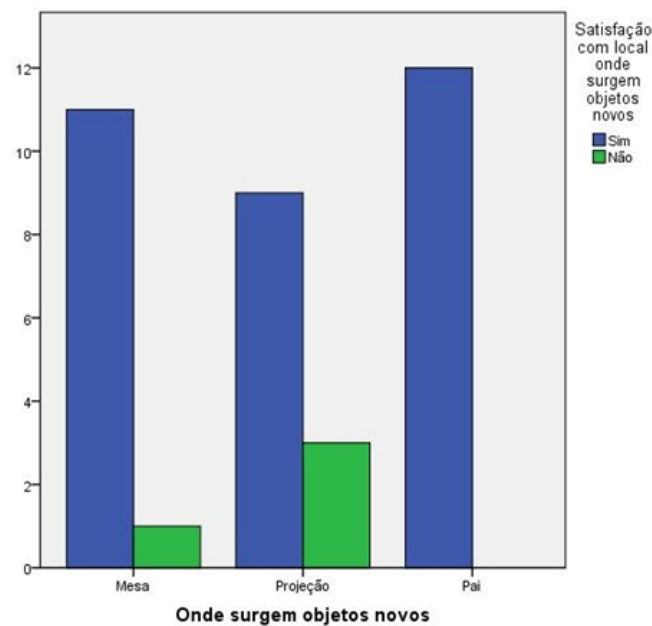


Figura 4.13: Estatística da satisfação dos participantes em relação do aparecimento dos objectos.

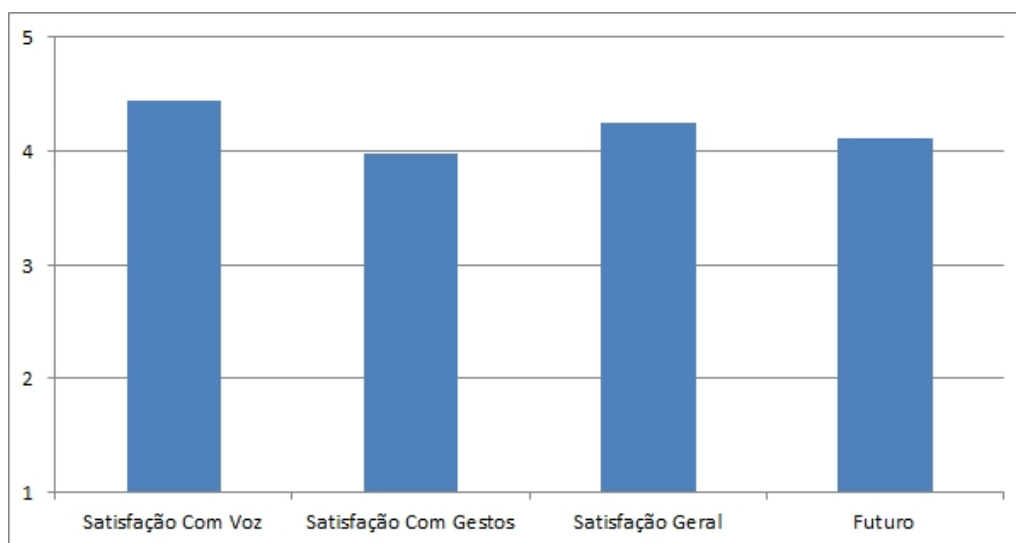


Figura 4.14: Estatística da satisfação dos participantes.

Por último pediu-se aos participantes para darem a sua opinião sobre a junção das duas superfícies de interação e utilizando as três modalidades de interação, dizendo se encontravam ou não vantagens em utilizá-las. Através do resultado que o gráfico 4.15 apresenta, é possível observar que a maioria dos participantes encontrou vantagens em juntar estas duas superfícies de interação, o que é um bom sinal para este trabalho, pois os participantes mostraram agrado e satisfação em conhecer este novo tipo de aplicações multimodais, multiutilizadores e multisuperfícies.

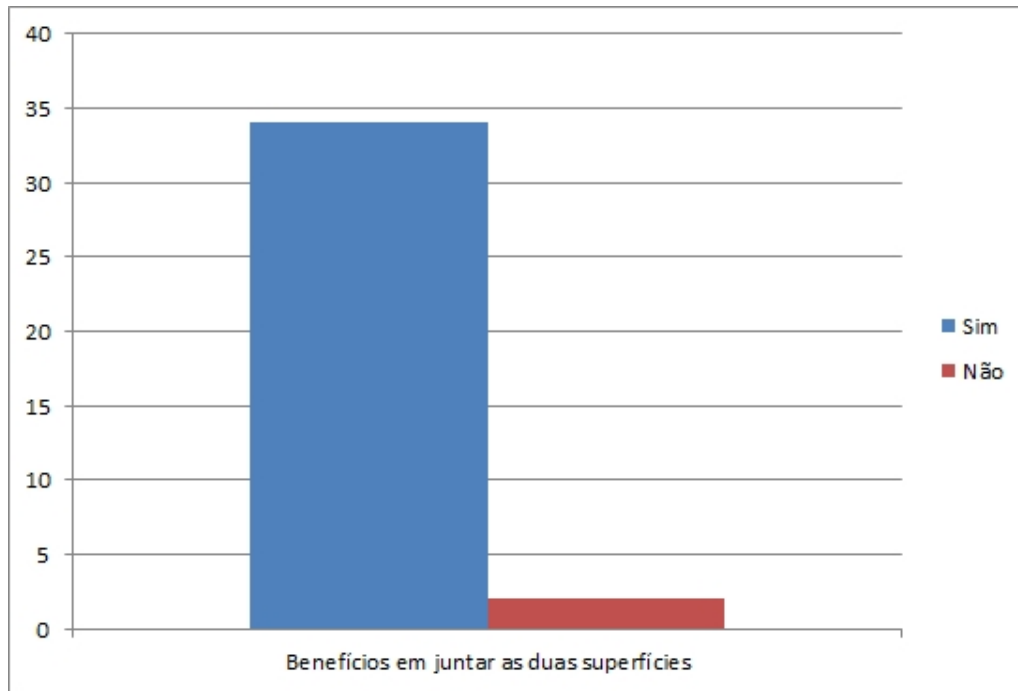


Figura 4.15: Estatística da satisfação dos participantes acerca do uso da diferentes superfícies.

4.2.6 Discussão dos resultados

Com a realização deste estudo conseguiu-se analisar o impacto das três variáveis propostas e também outras variáveis de satisfação dos participantes. Em relação ao número de pessoas em torno da mesa, foi observado que quanto maior o número de pessoas, maior é o número de comandos executados, sendo que a maior diferença é notada entre os grupos de uma pessoa e os grupos de três pessoas. O número de pessoas em torno da aplicação também causa impacto nas modalidades usadas, pois o número de comandos de seleção de botão, o número de comandos por gesto e o número de comandos por voz aumentam com o aumento do número de participantes, tendo mais impacto nos comandos dados através de gestos, ou seja, o grupo com uma pessoa executava menos gestos do que grupo de três pessoas. Assim entende-se que em aplicações deste tipo a interação aumente com o aumento do número de pessoas, mesmo que as tarefas pedidas sejam exatamente as mesmas e possam ser executadas da mesma forma e com o mesmo número de comandos.

Quando se estuda a abertura das cópias nas superfícies fica bem claro de que estas não podem aparecer na superfície contrária ao seu original, pois este procedimento não é do agrado dos utilizadores. Daqui tiramos a conclusão que em aplicações deste tipo, que tenham a ação de copiar, é muito recomendado que as cópias abram na mesma superfície que o original. Este resultado também é reforçado pelas observações feitas no momento dos testes. Os participantes mostravam-se descontentes quando a cópia aparecia na superfície contrária à que eles estavam a trabalhar pois perdiam um pouco o fluxo

de trabalho, porque tinham que ir puxar a cópia para a superfície que estavam a trabalhar e então depois voltar a executar a tarefa.

Em relação ao aparecimento de novos objetos nas superfícies não foi identificado nenhum impacto na aplicação sobre o número de comandos dados e a modalidade usada para fazer as ações pretendidas. Mas em termos de satisfação dos participantes pode-se concluir que a melhor opção neste tipo de aplicações multimodais, multiutilizador e multisuperfície é fazer aparecer o objeto na superfície em que o utilizador está a trabalhar no momento, ou seja, colocar os objetos a aparecer na superfície onde o seu “Pai” ou antecessor se encontra. Durante os testes também foi observado que quando o objeto aparecia sempre na projeção o participante puxava esse mesmo objeto para a mesa interativa, para poder tê-lo mais perto de si. Talvez este seja o fator causador da falta de impacto desta variável.

Este estudo também mostrou que os gestos e os comandos de voz aplicados nesta aplicação podem servir de referência a aplicações deste tipo pois a satisfação dos participantes foi bastante positiva no que toca aos gestos e comandos de voz permitidos. Esta observação vem validar, de certa forma, o processo de desenho apresentado no capítulo 3, visto que os gestos e comandos de voz usados foram identificados com o primeiro protótipo e os participantes, sendo outras pessoas, mostraram-se satisfeitos com eles.

Durante os testes realizados foi também observado se as ações permitidas pela aplicação faziam sentido ou não. Constatou-se que sim, pois os participantes minimizavam e maximizavam os objetos para ganhar mais espaço na mesa sem ter que os fechar. Também a ação de limpar a mesa para que ficasse desimpedida e com espaço foi utilizada, em situações que os participantes tinham imensa informação mas queriam tirar aquilo tudo dali de uma forma rápida e eficiente. A opção de copiar também foi bastante usada principalmente nos grupos de 3 pessoas pois cada um queria trabalhar com o seu objeto. A função de trocar objetos não poderia ser mais adequada a este tipo de aplicações porque como existem duas superfícies, faz todo o sentido haver essa troca de informação. A função de sincronizar teve uma aplicação curiosa pelos participantes, pois estes usaram-na para manipular os objetos na projeção através da mesa interativa, ou seja, o objeto era sincronizado com outro igual (cópia) e depois enviado para a projeção. De seguida era alterado o objeto na mesa e o objeto da projeção também sofria a alteração. Esta ação foi utilizada também para a marcação de voos nos grupos de 3 pessoas. Logo, esta ação de sincronizar objetos também tem sentido em estar presente neste tipo de aplicações, fazendo mais sentido, principalmente em cenários multiutilizador.

Outro dado bastante importante foi a opinião dos participantes sobre os benefícios da junção destas superfícies. A maioria concordou que era vantajoso utilizar aplicações deste tipo e alguns até fizeram comentários de que era melhor para partilhar informação e que poderia ser aplicado em apresentações de trabalhos, em consultórios médicos, reuniões de empresa, formações e outras tantas sugestões que foram dadas.

Capítulo 5

Conclusão

Neste capítulo apresentam-se as conclusões obtidas depois dos estudos realizados e abrangem-se perspectivas para trabalho futuro.

5.1 Conclusão

Durante a realização deste trabalho foram exploradas formas de melhorar a usabilidade em interação gestual com ou sem voz, em múltiplas superfícies, uma de toque direto e outra superfície onde não existe qualquer contacto direto entre os periféricos de entrada e os utilizadores. Com este fim realizaram-se dois estudos de modo a contribuir para o melhoramento do modo de desenvolvimento de desenhos de aplicações deste tipo.

A primeira impressão que se consegue obter com este trabalho é a adequação do processo de desenho, aqui apresentado, a aplicações multimodais, multiutilizador e multisuperfície. O primeiro estudo não teve a parte de multiutilizador, que foi adicionado mais tarde, mas os gestos e comandos de voz compilados na primeira etapa do processo, foram bem aceites pelos participantes no segundo estudo efetuado, o que leva à conclusão que, este processo pode servir para orientar o desenho deste tipo de aplicações.

No primeiro estudo, através da técnica do Feiticeiro de Oz, de um protótipo e de um conjunto de ações, foi pedido aos participantes para executarem as tarefas dadas, com o objetivo de encontrar gestos e comandos de voz padrão para cada ação. Os gestos e comandos de voz foram encontrados para todas as ações. No entanto não foi possível para determinadas ações encontrar apenas só um gesto, como é caso dos gestos para as ações copiar e sincronizar. Quando era utilizada a modalidade de comandos de voz juntamente com os gestos as ações que tinham gerado dúvida, como o copiar, deixaram de o fazer. Com a utilização da voz a manipulação dos objetos torna-se mais intuitiva, pois passa-se a conseguir uma manipulação direta dos objetos. Deste modo é possível concluir que para ações que sejam menos intuitivas e mais abstratas, é uma vantagem suportar interação por voz pois assim permite-nos ultrapassar o problema.

Os comandos de voz foram os mais preferidos para a superfície de projeção. Como

não existia toque direto na superfície, alguns dos participantes não sabiam o que haviam de fazer e um deles sugeriu a utilização de voz, mesmo sem saber que ser-lhe-ia pedido para a usar no próximo passo do teste.

A voz é um fator muito importante neste tipo de aplicações pois permite solucionar certas ambiguidades que surgiram, como na ação de copiar em que alguns participantes simulavam o gesto de fazer zoom in. Assim, com todas estas observações sobre a utilização de comandos de voz, é altamente recomendado a utilização de comandos de voz em aplicações que permitam multimodalidade e multisuperfície.

Esta sessão experimental serviu também para saber se o design escolhido para representar visualmente as ações pretendidas fazia algum sentido para os participantes. Assim pediu-se a opinião dos participantes para o reconhecimento daquela ação e depois a sua opinião sobre melhorias e alterações a aplicar. Observou-se que os botões de copiar e sincronizar eram aqueles que causavam mais confusão aos participantes.

A segunda sessão experimental surgiu com o objetivo de avaliar e estudar os fatores que influenciam o uso deste tipo de aplicação. Também serviu para avaliar os gestos e os comandos de voz que foram recolhidos na sessão anterior. Verificou-se que os gestos estavam adequados à ação pretendida uma vez em nenhum dos testes os participantes sentiram dificuldade em fazer o gesto ou em dizer o comando de voz. Esta sessão foi executada sem a técnica do Feiticeiro de Oz uma vez que o protótipo estava preparado para reconhecer os gestos e comandos de voz dados pelos utilizadores, baseados, como já foi referido anteriormente, no primeiro estudo.

Em aplicações que suportem ações como as que foram apresentadas neste trabalho, é recomendado a utilização do desenho utilizado neste segundo protótipo, pois os participantes não tiveram qualquer dificuldade em perceber a que ação correspondia determinado botão.

Foi analisado o número de pessoas em torno da aplicação. O resultado mostra que este fator altera a interação, aumentando o número de comandos efetuados, entre os grupos de trabalho. O que não se sabe, é se este aumento da interação é vantajoso ou não para a produtividade de cada pessoa inserida no grupo. No entanto, ao nível do grupo, foi possível verificar que todas as tarefas foram completadas com sucesso independentemente do número de pessoas no grupo. Grupos maiores executaram mais comandos para as mesmas tarefas (o que resulta numa menor eficiência), mas em resultados semelhantes ao nível da eficácia e da satisfação. O número de pessoas também tem um impacto nas modalidades de interação. Existiam três modalidades: comandos por seleção por botão; comandos por gestos; e comandos por voz. A conclusão a que se chegou foi que um grupo de uma pessoa tinha menos interação gestual do que um grupo de três pessoas. Uma outra variável que se estudou foi a abertura das cópias em função da funcionalidade da aplicação, ou seja, se as cópias abriam na mesma superfície que o original ou se abriam na superfície contrária. Os resultados permitem afirmar que os participantes não se sentem

confortáveis quando as cópias abrem na superfície contrária. Assim, na produção deste tipo de aplicações, que tenham a funcionalidade de copiar, é recomendado que as cópias dos objetos apareçam na mesma superfície que o original.

Outro aspecto analisado foi o aparecimento dos objetos nas diferentes superfícies. Esta foi das variáveis que não causou tanta diferença dos resultados, tendo a maioria dos participantes achado normal qualquer dos três comportamentos: abrir tudo na mesa, abrir tudo na projeção e abrir onde estava o seu objeto antecessor. Esta última foi a única forma que não levou uma nota negativa de qualquer participante. Portanto, para aplicações multimodais, multiutilizador e multisuperfície, recomenda-se que os objetos apareçam na superfície onde o objeto antecessor se encontra, até porque se observou que o comportamento típico dos participantes era copiar o objeto para a superfície onde já se encontravam a trabalhar anteriormente (a superfície do objeto "pai").

Em relação à opinião geral dos participantes pode-se concluir que avaliaram de forma positiva a interação aqui apresentada.

Em suma, existem recomendações importantes, a ter em conta na produção de aplicações multimodais, multiutilizador e multisuperfície. Em primeiro lugar temos o uso da voz. Sem os comandos de voz, os utilizadores não sabem fazer determinadas ações, por isso é indispensável existir esta interação. Em relação às cópias, chegou-se à conclusão que as cópias devem sempre abrir na superfície onde está o objeto original. Outra recomendação é deixar o utilizador escolher a superfície que quer trabalhar, não obrigando a usar uma específica, pois isso causará frustração. Finalmente, os novos objetos devem aparecer na superfície onde se encontra o seu antecessor.

5.2 Trabalho Futuro

A interação gestual é hoje em dia um meio comum de entrada de dados e está na contida nas mais variadas plataformas, como os tablets e smartphones.

Este trabalho focou-se na interação gestual com a junção de comandos de voz em múltiplas superfícies e no desenho de aplicações multimodais, multiutilizador e multisuperfície. O próximo passo seria melhorar a interface da aplicação tendo em conta os resultados obtidos através do estudo efetuado. Seria também interessante melhorar a interação gestual sem toque quando não é possível a utilização de voz, como em ambientes com muito ruído ou onde a voz não seja socialmente aceitável.

Como este tipo de interação e aplicações ainda não foram muito estudadas, mas estão em voga, seria importante formalizar um conjunto de recomendações que orientassem os possíveis interessados, na criação de uma aplicação que explore este tipo de interação, tendo este trabalho contribuído para iniciar esse processo.

Apêndice A

Questionário

Informações sobre o participante:

Idade:

Sexo:

Experiência com superfícies sensíveis ao toque?
Sim ou Não?

Experiência com dispositivos de reconhecimento corporal (como o Kinect)?
Sim ou Não?

Experiência com dispositivos de interação por voz?
Sim ou Não?

Grupo I

De uma escala de 1 a 5, classifica as seguintes afirmações:

Q1: Os comandos de voz são apropriados para realizar as acções pretendidas pelo utilizador.

Não Concordo 1 2 3 4 5 Concordo plenamente

Q2: Os gestos correspondem bem às expectativas.

Não Concordo 1 2 3 4 5 Concordo plenamente

Q3: Foi uma experiência agradável utilizar este tipo de interação.

Não Concordo 1 2 3 4 5 Concordo plenamente

Q4: Futuramente, gostaria de utilizar este tipo de tecnologia para escolher as férias com a família ou amigos.

Não Concordo 1 2 3 4 5 Concordo plenamente

Grupo II

Respostas SIM ou NÃO, com justificação

Q5: As cópias que foram feitas abriam na mesma superfície que o original. Esta acção pareceu-te normal?

SIM/NÃO? Se NÃO, porquê?

Q6: Quando um menu era chamado ele aparecia na mesa interactiva Esta acção pareceu-te normal?

SIM/NÃO? Se NÃO, porquê?

Q7: Achas que a junção desta projecção à mesa interativa traz benefícios?

SIM/NÃO? Se SIM, quais? Se NÃO, porquê?

Tens sugestões? (exemplo: Em que outras situações usarias este tipo de interação?)

Bibliografia

- [1] Community core vision. [online:]. <http://ccv.nuigroup.com/>.
- [2] Dsi - diffused surface illumination. <http://iad.projects.zhdk.ch/multitouch/?p=90>.
- [3] Kinect for windows sensor components and specifications. <http://msdn.microsoft.com/en-us/library/jj131033.aspx>.
- [4] Nodejs. [online:]. <http://nodejs.org/>.
- [5] Tuio org. [online:]. <http://www.tuio.org/>.
- [6] Mark Billinghurst. Gesture based interaction. *Continuum*, 31(1):1–35, 2011.
- [7] Richard A. Bolt. ‘put-that-there’: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, SIGGRAPH ’80, pages 262–270, New York, NY, USA, 1980. ACM.
- [8] Marie-Luce Bourguet and Akio Ando. Synchronization of speech and hand gestures during multimodal human-computer interaction. In *CHI 98 conference summary on Human factors in computing systems*, CHI ’98, pages 241–242, New York, NY, USA, 1998. ACM.
- [9] Hélène Cochet and Jacques Vauclair. Pointing gesture in young children: Hand preference and language development. *Gesture*, 10(2):129–149, 2010.
- [10] Andrew Cunningham, Ben Close, Bruce Thomas, and Peter Hutterer. Design and impressions of a multi-user tabletop interaction device. In *Proceedings of the Eleventh Australasian Conference on User Interface - Volume 106*, AUIC ’10, pages 71–79, Darlinghurst, Australia, Australia, 2010. Australian Computer Society, Inc.
- [11] Chi Tai Dang, Martin Straub, and Elisabeth André. Hand distinction for multi-touch tabletop interaction. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS ’09, pages 101–108, New York, NY, USA, 2009. ACM.

- [12] Carlos Duarte and António Neto. Gesture interaction in cooperation scenarios. In *Proceedings of the 15th international conference on Groupware: design, implementation, and use*, CRIWG'09, pages 190–205, Berlin, Heidelberg, 2009. Springer-Verlag.
- [13] Jacob Eisenstein and Randall Davis. Visual and linguistic information in gesture classification. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.
- [14] Julien Epps, Serge Lichman, and Mike Wu. A study of hand shape use in tabletop gesture interaction. In *CHI '06 extended abstracts on Human factors in computing systems*, CHI EA '06, pages 748–753, New York, NY, USA, 2006. ACM.
- [15] Tiago Uttini Gomes. ESTUDO DE MODOS DE COMANDO EM CENÁRIOS DE INTERACÇÃO GESTUAL. Master's thesis, Faculdade de Ciências da Universidade de Lisboa, 2010.
- [16] Carl Gutwin and Saul Greenberg. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, CSCW '98, pages 207–216, New York, NY, USA, 1998. ACM.
- [17] Uta Hinrichs and Sheelagh Carpendale. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 3023–3032, New York, NY, USA, 2011. ACM.
- [18] X. Lu, M. Unoki, and S. Nakamura. Normalization on the modulation spectrum of the subband temporal envelopes for automatic speech recognition in reverberant environments. In *Proceedings of the 3rd International Universal Communication Symposium*, IUCS '09, pages 247–254, New York, NY, USA, 2009. ACM.
- [19] Wolfgang M'ahr, Richard Carlsson, Jonas Fredriksson, Olivier Maul, and Morten Fjeld. Tabletop interaction: research alert. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, NordiCHI '06, pages 499–500, New York, NY, USA, 2006. ACM.
- [20] Shahzad Malik, Abhishek Ranjan, and Ravin Balakrishnan. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, UIST '05, pages 43–52, New York, NY, USA, 2005. ACM.
- [21] Audrey Mbogho and Michelle Katz. The impact of accents on automatic recognition of south african english speech: a preliminary investigation. In *Proceedings of*

- the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, SAICSIT '10*, pages 187–192, New York, NY, USA, 2010. ACM.
- [22] Microsoft. Common commands in speech recognition. <http://windows.microsoft.com/en-US/windows7/Common-commands-in-Speech-Recognition>.
- [23] Madoka Miki, Chiyomi Miyajima, Takanori Nishino, Norihide Kitaoka, and Kazuya Takeda. An integrative recognition method for speech and gestures. In *Proceedings of the 10th international conference on Multimodal interfaces, ICMI '08*, pages 93–96, New York, NY, USA, 2008. ACM.
- [24] Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010, GI '10*, pages 261–268, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.
- [25] Christian Müller-Tomfelde, Kelvin Cheng, and Jane Li. Pseudo-direct touch: interaction for collaboration in large and high-resolution displays environments. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference, OzCHI '11*, pages 225–228, New York, NY, USA, 2011. ACM.
- [26] Joana Neca and Carlos Duarte. Evaluation of gestural interaction with and without voice commands. In *Proceedings of IHCI 2011 – IADIS International Conference Interfaces and Human Computer Interaction 2011, Rome, Italy, 2011*.
- [27] Joana Raimundo Neca. INTERAÇÃO GESTUAL SEM SUPERFÍCIES DE APOIO. Master's thesis, Faculdade de Ciências da Universidade de Lisboa, 2011.
- [28] C L Nehaniv. Classifying types of gesture and inferring intent. *Robotics*, page 74, 2005.
- [29] António Regueiras Neto. INTERACÇÃO POR TOQUE EM MÚLTIPLAS SUPERFÍCIES. Master's thesis, Faculdade de Ciências da Universidade de Lisboa, 2009.
- [30] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. Multimodal human discourse: gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9:171–193, September 2002.
- [31] Miguel Sales Dias, Rafael Bastos, João Fernandes, João Tavares, and Pedro Santos. Gesture-based human-computer interaction and simulation. chapter Using Hand

- Gesture and Speech in a Multimodal Augmented Reality Environment, pages 175–180. Springer-Verlag, Berlin, Heidelberg, 2009.
- [32] Huda Sarfraz, Sarmad Hussain, Riffat Bokhari, Agha Ali Raza, Inam Ullah, Zahid Sarfraz, Sophia Pervez, Asad Mustafa, Iqra Javed, and Rahila Parveen. Large vocabulary continuous speech recognition for urdu. In *Proceedings of the 8th International Conference on Frontiers of Information Technology*, FIT '10, pages 1:1–1:5, New York, NY, USA, 2010. ACM.
- [33] Orit Shaer, Megan Strait, Consuelo Valdes, Taili Feng, Michael Lintz, and Heidi Wang. Enhancing genomic learning through tabletop interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2817–2826, New York, NY, USA, 2011. ACM.
- [34] Mara G. Silva and Doug A. Bowman. Body-based interaction for desktop games. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, CHI EA '09, pages 4249–4254, New York, NY, USA, 2009. ACM.
- [35] JaYoung Sung, Rebecca E. Grinter, and Henrik I. Christensen. "pimp my roomba": designing for personalization. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 193–196, New York, NY, USA, 2009. ACM.
- [36] Desney S. Tan, Darren Gergle, Peter Scupelli, and Randy Pausch. With similar visual angles, larger displays improve spatial performance. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 217–224, New York, NY, USA, 2003. ACM.
- [37] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 1181–1190, New York, NY, USA, 2006. ACM.
- [38] Edward Tse, Saul Greenberg, and Chia Shen. Gsi demo: multiuser gesture/speech interaction over digital tables by wrapping single user applications. In *Proceedings of the 8th international conference on Multimodal interfaces*, ICMI '06, pages 76–83, New York, NY, USA, 2006. ACM.
- [39] Edward Tse, Chia Shen, Saul Greenberg, and Clifton Forlines. Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '06, pages 336–343, New York, NY, USA, 2006. ACM.

- [40] Stephen Volda, Matthew Tobiasz, Julie Stromer, Petra Isenberg, and Sheelagh Carpendale. Getting practical with interactive tabletop displays: designing for dense data, "fat fingers," diverse interactions, and face-to-face collaboration. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 109–116, New York, NY, USA, 2009. ACM.
- [41] Andrew D. Wilson and Aaron F. Bobick. Real-time online adaptive gesture recognition. In *Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, RATFG-RTS '99, pages 111–, Washington, DC, USA, 1999. IEEE Computer Society.
- [42] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 1083–1092, New York, NY, USA, 2009. ACM.
- [43] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07, pages 159–168, New York, NY, USA, 2007. ACM.
- [44] Mike Wu and Ravin Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, UIST '03, pages 193–202, New York, NY, USA, 2003. ACM.
- [45] Ying Yin and Randall Davis. Toward natural interaction in the real world: real-time gesture recognition. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, pages 15:1–15:8, New York, NY, USA, 2010. ACM.

